The Computational Neurobiology of Perceptual Awareness

Christopher Whyte

A thesis presented for the degree of
Doctor of Philosophy

January 2026

Faculty of Medicine and Health
The University of Sydney

Supervisor: Prof. James M. Shine
Co-Supervisors: Dr. Brandon R. Munn & Dr. Eli J. Müller

To my parents, Lynda and Paul, for giving me the space, love, and support that I needed to figure out, or perhaps discover, the type of person I want to be.

And to my supervisor and friend, Mac Shine, for being the type of scientist and mentor I hope to one day become.

Abstract

Understanding the neurobiological basis of consciousness faces a critical obstacle: the explanatory gap between human and animal model-based neuroscience. Human participants can perform a rich array of behavioural tasks that precisely target psychological constructs, but our ability to non-invasively record and control neural activity is limited. Conversely, in animal models, it is possible to invasively record and perturb neural activity with exquisite detail and precision, however the range of psychological constructs that can be studied is highly restricted. This thesis bridges this explanatory divide by constructing biophysical models based on nonlinear dynamical systems theory that explicitly link systems neuroscience with human psychophysics. Part one focuses on the neurobiology of perceptual awareness. Following a review of the state of the art in systems neuroscience, which identifies an essential role for thalamocortical loops in controlling the state and contents of consciousness, I introduce a spiking neural network model of perceptual awareness that incorporates the key cellular elements of thalamocortical loops covered in the review. The model reproduces neural signatures of perceptual awareness found in mouse models and generalises to visual rivalry, generating a series of novel predictions. Part two abstracts away from cellular detail to derive law-like expressions for awareness thresholds in a novel paradigm: tracking continuous flash suppression (tCFS). I demonstrate that a reduced neural mass model can be derived from the thalamocortical dynamics established in part one. Simulations of the reduced model confirm that the mechanisms governing binocular rivalry generalise to tCFS. I then derive closed-form expressions for dominance durations and awareness thresholds, validating them against existing psychophysical data. Collectively, the thesis provides a quantitative roadmap for integrating the tools of systems neuroscience with human psychophysics.

Statement of Originality

I hereby declare that the work presented in this thesis, titled 'The Computational Neurobiology of Perceptual Awareness', submitted to The University of Sydney in fulfilment of the requirements for the award of the degree of Doctor of Philosophy, is a genuine record of the research work carried out under the supervision of Professor James M. Shine. This is to certify that the content of this thesis is my own work. This thesis has not been submitted for any other degree or purpose. I certify that the intellectual content of this thesis is the product of my own work, and that all assistance received in preparing this thesis, and all sources, have been acknowledged.

Christopher Whyte

Certificate of Originality

As the primary supervisor for the candidature upon which this thesis is based, I hereby certify that this thesis is sufficiently well presented for examination and that it adheres to the prescribed word limit. I confirm that the Author's declaration above is correct and that the authorship attribution statements below are correct.

Professor of Systems Neurobiology James M. Shine

Author Attribution Statement

We, the undersigned, acknowledge the following author attributions:

This thesis represents the work of Christopher Whyte. The primary supervisor, Professor James M. Shine, co-supervisor Dr Eli J. Müller, and auxiliary supervisor Dr Brandon R. Munn provided guidance across all studies conducted as a part of the thesis research and were involved in reviewing the introduction and discussion of this thesis.

The work presented in this thesis was performed as part of a collaborative research program, with contributions from all authors summarised below.

Chapter 1 of this thesis is partially (sect 1.1) based on the published paper: **Whyte, C. J.,** Corcoran, A. W., Robinson, J., Smith, R., Moran, R. J., Parr, T., ... & Hohwy, J. (2025). On the minimal theory of consciousness implicit in active inference. Physics of Life Reviews.

Contributions: Christopher Whyte conceptualised the paper, wrote the original draft, and edited the manuscript. Andrew Corcoran, Jonathan Robinson, Ryan Smith, Rosalyn J. Moran, Thomas Parr, Karl J. Friston, Anil K. Seth, and Jakob Hohwy provided feedback and conceptual guidance on the subset of the paper included in the thesis.

Chapter 2 of this thesis is based on the published paper: **Whyte, C. J.,** Redinbaugh, M. J., Shine, J. M., & Saalmann, Y. B. (2024). Thalamic contributions to the state and contents of consciousness. Neuron, 112(10), 1611-1625.

Contributions: Christopher Whyte conceptualised, wrote, and edited the review in collaboration with Michelle J. Redinbaugh, James M. Shine, and Yuri B. Saalmann. All authors contributed equally.

Chapter 3 of the thesis is based on the published paper: **Whyte, C. J.,** Müller, E. J., Aru, J., Larkum, M., John, Y., Munn, B. R., & Shine, J. M. (2025). A burst-dependent thalamocortical substrate for perceptual awareness. PLOS Computational Biology, 21(4), e1012951.

Contributions: Christopher Whyte conceptualised and implemented the model, wrote the first draft, and edited the manuscript. James M. Shine provided conceptual guidance and supervision on the design of the model and edited the manuscript. Eli J. Müller provided technical supervision and edited the manuscript. Brandon R. Munn provided technical supervision, aided in the implementation of the model, and edited the manuscript. Jaan Aru, Yohan John, and Matthew Larkum edited the manuscript and provided conceptual guidance

Chapter 4 of the thesis is not based upon an existing publication and is designed to serve as a bridge between the models implemented in part 1 and 2 of the thesis.

Contributions: Christopher Whyte wrote the chapter and designed and implemented the

model. James M. Shine, Eli J. Müller, and Brandon R. Munn provided editorial input.

Chapter 5 of the thesis is based upon the preprint that is currently under review: **Whyte, C. J.,** Wilson, H. R., Shine, J. M., & Alais, D. (2025). A Minimal Physiological Model of Perceptual Suppression and Breakthrough in Visual Rivalry. bioRxiv, 2025-05.

Contributions: Christopher Whyte conceptualised and implemented the model, wrote the first draft, and edited the manuscript. Hugh R. Wilson aided in the design of the model and edited the manuscript. James M. Shine edited the manuscript and provided conceptual guidance. David Alais contributed to the first draft, and edited the manuscript.

Chapter 6 of the thesis is based upon the preprint that is currently under review: **Whyte, C. J.,** Wilson, H. R., Tobin, S., Munn, B. R., Safavi, S., Muller, E. J., ... & Alais, D. (2025). A Minimal Quantitative Model of Perceptual Suppression and Breakthrough in Visual Rivalry. arXiv preprint arXiv:2510.17154.

Contributions: Christopher Whyte conceptualised and implemented the model, performed the mathematical derivations, analysed the data, wrote the first draft, and edited the manuscript. Hugh R. Wilson was involved in the conceptualisation of the model and edited the manuscript. Shay J. Tobin aided in (and checked) the mathematical derivations, and edited the manuscript. Brandon R. Munn, Shervin Safavi, Eli J. Müller, Jayson Jeganathan, and James M. Shine edited the manuscript and provided conceptual guidance on the development of the model and analysis of the data. Matt Davidson and David Alais collected the empirical data and edited the manuscript.

Chapter 7 of this thesis is partially based (sect 7.4) on the published paper: **Whyte, C. J.,** Redinbaugh, M. J., Shine, J. M., & Saalmann, Y. B. (2024). Thalamic contributions to the state and contents of consciousness. Neuron, 112(10), 1611-1625.

Contributions: Christopher Whyte conceptualised, wrote, and edited the relevant section in collaboration with Michelle J. Redinbaugh, James M. Shine, and Yuri B. Saalmann. All authors contributed equally.

Christopher Whyte
Candidate
27 January 2026

Professor James M Shine
Primary Supervisor
27 January 2026

Dr Eli J. Müller
Co-Supervisor
27 January 2026

# Additional publications authored during candidature (* denotes equal contribution)

2025

Bassler, M., Emming, L., **Whyte, C. J.,** Huis in't Veld, G., Suzuki, M., & Pennartz, C. (2025). Continuous Flash Suppression responses in mouse visual cortex: stimulus laterality and anesthesia effects. bioRxiv, 2025-11.

Whyte, C. J., Manohar, S. G., Feredoes, E., & Woolgar, A. (2025). A plastic attractor model of flexible rule-based selective attention. bioRxiv, 2025-09.

Rodriguez-Garcia, A., **Whyte, C. J.,** Munn, B. R., Mei, J., Shine, J. M., & Ramaswamy, S. (2025). The role of gain neuromodulation in layer-5 pyramidal neurons. arXiv preprint arXiv:2507.03222.

Tan, J. B., Orlando, I. F., **Whyte, C.,** Bryant, A. G., Munn, B. R., Baracchini, G., ... & Shine, J. M. (2025). Cerebellar and subcortical contributions to working memory manipulation. Communications Biology, 8(1), 1028.

Bryant, A. G., & **Whyte, C. J.** (2025). A data-driven approach to identifying and evaluating connectivity-based neural correlates of consciousness. bioRxiv, 2025-04.

Fisher, E. L., **Whyte, C. J.,** & Hohwy, J. (2025). An Active Inference Model of the Optimism Bias. Computational Psychiatry, 9(1), 3.

Wainstein, G*., **Whyte, C. J*.,** Martens, K. A. E., Müller, E. J., Medel, V., Anderson, B., ... & Shine, J. M. (2025). Evidence from pupillometry, fMRI, and RNN modelling shows that gain neuromodulation mediates task-relevant perceptual switches. eLife, 13, RP93191.

2024

Taylor, N. L., **Whyte, C. J.,** Munn, B. R., Chang, C., Lizier, J. T., Leopold, D. A., ... & Shine, J. M. (2024). Causal evidence for cholinergic stabilization of attractor landscape dynamics. Cell reports, 43(6).

2023

Munn, B. R., Müller, E. J., Aru, J., **Whyte, C. J.,** Gidon, A., Larkum, M. E., & Shine, J. M. (2023). A thalamocortical substrate for integrated information via critical synchronous bursting. Proceedings of the National Academy of Sciences, 120(46), e2308670120.

Artificial Intelligence Statement

During the preparation of the thesis, the Author (Christopher Whyte) used Grammarly and Gemini for grammar and text enhancement to improve the clarity of text. The Author confirms that where such modifications were suggested by these tools, the content was manually reviewed for accuracy and appropriateness before inclusion in the thesis.

During the preparation of the thesis, the Author used Gemini and GitHub Copilot (embedded within MATLAB and Visual Studio Code) to assist with MATLAB and Python code development in support of analysis and figure generation.

The author takes full responsibility for the submitted thesis and ensures the work is their own and has used generative AI within the parameters of use.

Australian Goverment Support

This research was supported by an Australian Government Research Training Program (RTP) Scholarship and a DVCR Strategic Postgraduate Scholarship awarded by the University of Sydney.

Acknowledgements

If one were to describe my personality in one word, it would likely be 'obsessive'. Probably because of this, I have always had an unusual relationship with education. I was a mediocre primary and high school student, was terribly bullied, and only barely managed to scrape my way into university, which I didn't attend until the age of 22. I had the slightly delusional idea that I wanted to be a professional BMX rider, and not much else seemed important. But something changed when I started university; my obsession shifted from BMX to philosophy, neuroscience, and maths. At university, I found a home. A small group of friends and lecturers, every bit as excited and passionate as I was. This thesis is a reflection of the nearly ten-year educational journey that I started as a 22-year-old undergraduate. Many of the faces have changed, but academia is every bit the intellectual and social home it always has been. I am forever indebted to the friends and mentors who have made it so.

First and foremost, I want to express my most heartfelt appreciation to my supervisor, Mac, who serves in my mind as a paradigmatic example of what a scientist ought to be: deeply and broadly curious, kind, pragmatic, infectiously excitable, and uncompromising when it matters. I cannot imagine a more stimulating and welcoming environment in which to pursue a PhD than the one you created.

To my undergraduate mentors, Richard Heersmink, the late Karola Stotz, Alex Gillett, Chris Hewitson, and David Kaplan, I extend my sincere thanks for fostering and guiding my endless and (I'm sure at points slightly tiresome) enthusiasm.

To my Masters (plural) advisers and mentors, Tom Carlson, Tijl Grootswagers, Amanda Robinson, and Alex Woolgar, I am incredibly grateful for the enormous amount of time each of you spent training me and for your kindness and patience with my (frustratingly) broad interests and often stochastic attention span.

To Ryan Smith, thank you for taking a chance on a random Masters student from Australia, taking me under your wing, spending endless hours teaching me how to build active inference models, and for continuing to be one of my greatest advocates.

To Jakob Hohwy, thank you for your sincere interest and encouragement, and for taking me and my work seriously from the first time I met you as a third-year undergraduate in a poster session, to today, as a collaborator, and friend. Had I told the undergraduate version of myself that I would write a paper with the author of *The Predictive Mind*, I would not have believed myself.

To my co-supervisors, Brandon and Eli, thank you for welcoming me into the group with such energy, interest, and encouragement, as well as for all the time both of you dedicated to training me. I wouldn't be the scientist I am today without you.

To the students and postdocs in the Shine lab and Systems Neuroscience and Complexity

Group: Tash, Giulia, Josh, Gabe, Bella, Jungwoo, Jayson, and Teresa. Thank you for creating the lively, warm, and welcoming culture that has made the past three years some of the best years of my life.

To my collaborators, Michelle Redinbaugh, Yuri Saalmann, Jaan Aru, Yohan John, Matthew Larkum, Hugh Wilson, David Alais, Matt Davidson, Shervin Safavi, Andrew Corcoran, Jonathan Robinson, Thomas Parr, Karl Friston, Anil Seth, Mathis Bassler, and Cyriel Pennartz: thank you for your intellectual generosity and guidance. I would like to thank Matthew Larkum in particular for welcoming me into his home and lab in Berlin.

To the other PIs in the Systems Neuroscience and Complexity group: Ben Fulcher, Joe Lizier, and Claire O'Callaghan. Thank you for creating the intellectual and social environment that has been my treasured home for the past three years, with special thanks to Claire for being a constant source of laughter, advice, and book recommendations.

To my Cambridge friends: Molly, Tim, and Rachel. Thank you for being such wonderful friends and people. Our movie nights and bowling trips are some of my most treasured memories. Leaving Cambridge was incredibly difficult, but it was made just a little bit easier knowing that we would still be friends no matter the distance.

To my rock climbing friends: Chris, Steph, Jarrod, and Tom. It is so rare to make new friends in your thirties, let alone such good friends. Thank you for the fun and the laughter.

To Erica, Matt, and Annie (Nolan), thank you for welcoming me into your lovely group of fiends and for being so kind and supportive through some truly difficult times.

To my three closest friends, Shay, Dayvis, and Beth. Shay, if I have any semblance of mathematical competence, it is in large part due to your tutelage and encouragement. Thank you for your kindness and generosity of spirit, for welcoming me into your wonderful group of friends, and for being one of the most interesting people I have ever had the good fortune to call my friend. To Dayvis, my oldest friend, my brother, we have come such a long way together. I am so proud of the people we have become. It wouldn't have been possible without you. To Beth, thank you for your unwavering optimism and positivity, and for being such a devoted and loyal friend. I am so lucky to have you in my life.

To Annie: thank you for your remarkable patience and understanding, and for reminding me what really matters.

Finally, to my parents, Lynda and Paul, thank you for trusting me to make my own mistakes, for the unending love and support, and for allowing me the space to figure out who it is that I want to be. To my dad, Paul, thank you for passing on your love of science and for making the world feel like a puzzle that, however difficult, always has an answer. To my mum, Lynda, thank you for your unconditional love and understanding, and for passing on your reverence for the natural world.

Contents

1 Towards a Computational Neurobiology of Consciousness **2**
1.1 Consciousness and its theories . . . 2
1.2 The strengths and limitations of systems neuroscience for consciousness research 4
1.3 Aims and scope: quantitative models are a precondition for effective theory . . 6
1.4 Essential elements for a computational neurobiology of perceptual awareness . 7
1.5 Thesis structure . . . 9
Chapter 1 references . . . 11

Part 1: The Thalamocortical Basis of Perceptual Awareness **15**

2 Thalamocortical Control of the State and Contents of Consciousness **17**
2.1 Chapter 2 introduction . . . 18
2.2 The anatomical basis of thalamocortical connectivity . . . 18
2.3 Thalamic contributions to conscious state . . . 20
2.3.1 Thalamocortical microcircuit contributions to conscious state . . . 20
2.3.2 Thalamocortical mesocircuit contributions to conscious state . . . 23
2.4 Thalamocortical contributions to conscious contents . . . 26
2.4.1 Thalamus gates conscious contents via control of perceptual thresholds 26
2.5 Chapter 2 discussion . . . 28
Chapter 2 references . . . 30

3 A Burst-dependent Thalamocortical Substrate for Perceptual Awareness **38**
3.1 Chapter 3 introduction . . . 39
3.2 Chapter 3 results . . . 40
3.2.1 A spiking corticothalamic model recreates key features of cellular physiology . . . 40
3.2.2 Thalamocortical model reproduces empirical signatures of threshold detection . . . 42
3.2.3 Thalamocortical spiking model generalises to visual rivalry . . . 44
3.2.4 Thalamocortical spiking model conforms to Levelt's propositions . . . 46
3.2.5 Generating testable predictions through *in silico* electrophysiology . . . 47
3.3 Chapter 3 discussion . . . 50
3.4 Chapter 3 methods . . . 53
3.4.1 Thalamocortical spiking neural network . . . 53
3.4.2 Distance to bifurcation . . . 57
3.4.3 Psychometric and neurometric functions . . . 57

3.5 Chapter 3 supplementary material . . . 58
3.5.1 Apical compartment phase plane . . . 58
3.5.2 Sweeping the magnitude of model perturbations . . . 58
3.5.3 Robustness of rivalry duration across burstiness parameters . . . 59
3.5.4 Dynamical regime underlying visual rivalry . . . 59
3.5.5 Key effects of visual rivalry simulations are preserved in scaled-up model 60
3.5.6 *In silico* electrophysiology supplemental figures . . . 62
Chapter 3 references . . . 63

Part 2: Dynamical Principles of Awareness and Suppression 67

4 From Cellular Neurobiology to Psychophysics 69
4.1 From thalamocortical dynamics to cortical masses: the case of binocular rivalry . . . 70
4.2 When do neurobiological details matter for psychology? . . . 72
4.3 Chapter 4 supplementary material . . . 74
4.3.1 Model parameters . . . 74
Chapter 4 references . . . 75

5 A Minimal Physiological Model of Perceptual Suppression and Breakthrough in Visual Rivalry 76
5.1 Chapter 5 Introduction . . . 76
5.2 Chapter 5 methods . . . 78
5.3 Chapter 5 results . . . 82
5.3.1 Suppression depth in binocular rivalry . . . 82
5.3.2 Suppression depth in tracking continuous flash suppression . . . 84
5.4 Chapter 5 discussion . . . 87
5.5 Chapter 5 supplementary material . . . 89
5.5.1 Model parameters . . . 89
5.5.2 Ideal observer readout . . . 89
Chapter 5 references . . . 91

6 A Minimal Quantitative Model of Perceptual Suppression and Breakthrough in Visual Rivalry 95
6.1 Introduction . . . 95
6.2 Chapter 6 methods . . . 97
6.3 Chapter 6 results . . . 101
6.3.1 Hysteresis and contrast rate . . . 101
6.3.2 Dominance duration and the depth of hysteresis . . . 104
6.3.3 Confirmation of novel prediction . . . 107
6.3.4 Additional predictions . . . 107
6.4 Discussion . . . 108
6.5 Chapter 6 supplementary material . . . 110
6.5.1 tCFS model parameters . . . 110
6.5.2 Derivation of dominance and suppression duration formula . . . 110
Chapter 6 references . . . 113

7 Thesis Conclusion 116
7.1 Where we have been and where are we going? . . . 116

7.2 When do neurobiological details matter? 119
7.3 How does the work align with existing theories of consciousness? 120
7.4 What might a mature thalamocortical theory of conscious contents look like? . 120
7.5 Concluding remarks 124
Chapter 7 references 125

Chapter 1

Towards a Computational Neurobiology of Consciousness

> The materials I use to answer the question — what is causality? — come from several disciplines including heavy reliance on neurobiology and nonlinear dynamics. In the words of computer technologists these two disciplines make up God's own firewall...
>
> — *Walter J. Freeman III*

1.1 Consciousness and its theories

Consciousness is a heterogeneous and multifaceted phenomenon. At first pass, the scientific study of consciousness can be divided into three related, but in practice largely independent, research programs: the study of contents, state, and self [1]. The contents of consciousness are the qualities or elements within experience that one is conscious of (e.g., the image of a red rose against a green background, or the aroma of freshly brewed coffee). Contents are studied by controlling for the physical attributes of a stimulus, and the overall state of consciousness (such as drowsiness), while varying the contents of subjective perception [2]. Conscious organisms also have different global states of consciousness, which are often assessed behaviourally (e.g., through the Glasgow Coma Scale; [3]) and are crucial to assess patients with disorders of consciousness. Such states include the vegetative state, various states of sleep, normal waking states, and perhaps states like delirium or psychedelia (for discussion, see [4]). Consciousness (at least in humans) is typically also accompanied by some form of minimal and/or narrative self-awareness [5] along with experiences of embodiment, selfhood, and personhood [6–8].

Most neuroscientific theories of consciousness take some subset of these phenomena as their explanatory target and construct a theory around the relevant empirical paradigms and findings [9]. For example, Global Workspace Theory [10, 11], and its contemporary incarnation, Global Neuronal Workspace Theory [12, 13], were constructed around the method of contrastive anal-

ysis, which treats the contents of consciousness as the independent variable varying whether a participant is conscious of some particular content evoked by a stimulus while holding all other variables constant (in so far as this is possible). The global state of consciousness was initially treated as a background condition for contents to become conscious [14], making Global Neuronal Workspace Theory chiefly a theory of conscious contents. Since then, the theory has also been expanded to explain experiments that manipulate the global state of consciousness through anaesthesia (for review see [13]). However, the initial assumption that conscious state is a background condition for contents, rather than a multidimensional construct that constrains the type of content that can become conscious, is arguably still present in contemporary versions of the theory (see [4, 15]). Higher-order Theories of consciousness [16–18] are, likewise, chiefly theories of conscious contents and are so far silent on the relationship between the contents and global state of consciousness. Similarly, Integrated Information Theory [19–21], another leading theory of consciousness, was developed with the explicit aim of solving the hard problem of consciousness (i.e., explaining why some physical structures generate subjective experience and others do not). Integrated information theory treats consciousness in wholly intrinsic terms, downplaying the role of overt behaviour, which arguably plays a major role in shaping not only what we are conscious of but also in determining the qualitative character of conscious contents [22, 23]. Other theories, such as the Self-model Theory of Subjectivity [24] or the Projective Consciousness Theory [25], focus on explaining the (seeming) presence of a self or first-person perspective. There are still other theories of consciousness beyond those cited above, many of which privilege specific explanatory targets and methodological approaches.

This state of affairs thus poses a double challenge: not only are theories of consciousness difficult to arbitrate between - on the basis of empirical evidence [26] - it is also sometimes not clear whether such theories aim to explain the same empirical data [9]. Still, there are at least two disciplines that all empirically productive neuroscientific theories of consciousness must draw evidence from: experimental psychology and neurobiology.

Experimental psychology is necessarily a part of any neuroscientific endeavour that aims to uncover the neural basis of a psychological function. The structure of a neural circuit may give hints to its function, but the nonlinearity and degeneracy present in even the simplest circuits make straight forward inference from structure to function effectively intractable [27]. We must therefore use carefully controlled experimental paradigms to generate separable conditions, defined in terms of behavioural accuracy or systematic subjective reports, which isolate the psychological construct of interest, allowing us to study neural circuitry in the operating regime we are trying to explain [28]. In the case of the neuroscience of consciousness, this typically involves contrasting conscious states with unconscious states. For example, in the case of conscious perception, sensations resulting from physically matched stimuli that are unconscious in one condition, but are consciously perceived in another, are sorted into conscious and unconscious conditions based on a participant's subjective report [2]. Or, when studying the global state of consciousness, anaesthetic-induced unconsciousness is typically contrasted with a restful waking state, where the difference between the two is operationalised in terms of some observable difference in behavioural responsiveness [2, 29].

The second source of evidence is, necessarily, neurobiological. Every major neuroscientific theory of consciousness aims to provide some kind of systematic explanatory bridge between neural mechanisms and some aspect of conscious experience. Indeed, many of the leading theories have neurobiological commitments across multiple scales and levels of analysis. For example, Global Neuronal Workspace Theory postulates that conscious perception occurs

when perceptual contents are "globally broadcast" by a large-scale fronto-parietal network via a nonlinear "ignition" event that makes information available across an otherwise modular network via long-range connections between layer 2 and 3 (L2/3) pyramidal neurons [12, 13, 30]. Local Recurrency Theory postulates that NMDA receptor dependent feedback between hierarchical levels of the visual hierarchy is sufficient for conscious vision [31]. Integrated Information Theory posits that conscious perception is dependent upon recurrent connectivity motifs within posterior cortex, leading to both macroscale and micro-scale predictions [19, 21]. Specifically, at the macro-scale, conscious perception is predicted to rely upon activity within the "posterior hot zone", and at the micro-scale the phenomenology of visual space has been proposed to rely upon the grid-like structure of V1 [32].

And yet, despite almost all of the major neuroscientific theories of consciousness having commitments across multiple scales and levels of analysis, the vast majority of existing empirical work has focused on the macro- and meso-scale predictions that can be studied in human neuroimaging. Here, however, high-resolution (i.e., layer-specific) recordings and precise causal manipulations (e.g., optogenetic and pharmacological manipulations) are exceedingly difficult and often impossible, restricting empirical progress and theoretical development to the level of large-scale networks. In contrast, the cellular and systems-level commitments of each theory are almost entirely unexplored theoretically, let alone tested empirically. Crucially, with technological advances allowing high throughput recordings of thousands of neurons simultaneously with techniques such as Neuropixels [33], and calcium imaging [34], and causal techniques, such as optogenetics [35] and Designer Receptors Exclusively Activated by Designer Drugs (DREADDs; [36]), if animals can be trained to perform tasks relevant to consciousness science, it is now possible to make causal inferences about the neural basis of consciousness [37, 38].

1.2 The strengths and limitations of systems neuroscience for consciousness research

Independent of the theoretically-driven work in the cognitive neuroscience of consciousness, great progress has recently been made in identifying the circuit-level mechanisms of conscious contents and conscious state using mouse models. Specifically, there is now good evidence that bursting in thick-tufted cortical layer five (L5) extratelencephalic ($L5_{ET}$) neurons, mediated by matrix-type projections in higher-order thalamus, plays a causal role in both conscious perception and anesthetic-induced unconsciousness [39–44]. This is theoretically significant: ET-type cells in cortical L5 have a unique morphology and connectivity that places them at the nexus of feedforward and feedback information flow in the cortex, and has them serve as the primary source of cortical input to subcortical structures [41].

The soma of these ET-type cells is located in L5b, receives feedforward inputs from "lower" cortical populations, and projects to subcortical structures including higher-order thalamic nuclei, superior colliculus, and striatum. The apical dendrites of $L5_{ET}$ cells are located in L1 and receive feedback projections from layer 5 of "higher" cortical areas and from matrix-type projections in higher-order thalamus [45–49]. Importantly, the apical dendrites of $L5_{ET}$ cells display nonlinear depolarisation dynamics - calcium plateau potentials - reminiscent of sodium spikes in the soma, but with much longer periods of depolarisation [50]. When a calcium plateau potential in the apical dendrites and a sodium spike in the soma co-occur (within 25-30 ms), $L5_{ET}$ cells exhibit a dramatic change in dynamical regime, switching from a regular

spiking to a bursting mode [51–54]. The switch from regular spiking to bursting depends on reliable communication between the apical dendrites and the soma which is controlled by matrix-type thalamic projections onto metabotropic glutamate and acetylcholine receptors on the oblique dendrites - or apical coupling zone - of $L5_{ET}$ cells [42].

In the context of simple tactile threshold-detection tasks, bursting in $L5_{ET}$ neurons in barrel cortex has been shown to distinguish hits and false alarms from misses and correct rejections [43, 44]. Crucially, this is not just a correlation; the threshold for perception can be causally controlled by manipulating the excitability of $L5_{ET}$ apical dendrites, either directly by manipulating the apical dendrites themselves, or via targeting downstream subcortical targets of $L5_{ET}$ neurons which project (either directly or indirectly) back to the apical dendrites. Optogenetic activation of $L5_{ET}$ apical dendrites decreases the animal's perceptual threshold, increasing both the hit and false alarm rates, and pharmacological inhibition of apical dendrites through the application of baclofen (a $GABA_B$ agonist) increases the animal's perceptual threshold [43]. Further, targeted inhibition of POm, a higher-order thalamic nuclei rich in matrix-type projections, superior colliculus, and the striatum, all of which project either directly or indirectly back to the apical dendrites of $L5_{ET}$, increases the animal's perceptual threshold [44]. There is, therefore, good reason to believe that the $L5_{ET}$ - higher-order thalamus loop is causally involved in conscious perception.

Perhaps even more remarkably, administration of a variety of general anaesthetics with distinct molecular mechanisms - including ketamine, isoflurane, and propofol - removes the influence of calcium plateau potentials on the somatic firing rate of $L5_{ET}$ cells [42], suggesting that the coupling between the two compartments may be a general mechanism related to the control of the global state of consciousness [39]. Indeed, in non-human primates, stimulation of matrix-rich higher-order thalamic nuclei, which project to both the apical dendrites and the apical coupling zone of $L5_{ET}$ neurons, restores consciousness even while the animal is actively under the inhibitory pressure of anaesthesia [55, 56]. This set of experimental results, combined with the aforementioned anatomical considerations, has led to the introduction of dendritic integration theory [41], which proposes that the coupling between apical and somatic compartments of $L5_{ET}$ cells controls both the global state of consciousness and the contents of consciousness. Specifically, bursting in $L5_{ET}$ cells acts as a gating mechanism that decides whether activity can freely propagate through corticocortical and thalamocortical loops or whether activity will remain isolated within a cortical column. Despite the title, dendritic integration theory is better thought of as an explanatory framework or working hypothesis, based upon a collection of systematic observations, rather than a full-fledged theory of consciousness.

Amidst the remarkable progress in the systems neuroscience of consciousness, a remaining limitation of the explanatory machinery of dendritic integration theory - and, for that matter, any explanatory framework based upon work in animal models - is that it is unclear whether (and to what extent) the circuitry identified in the animal model will generalise to the tasks studied in human and non-human primates, which often require elaborate verbal instructions or months of training, respectively [57]. The simplicity of the tasks and species-specific differences in neural architecture mean that it is far from certain whether the findings will generalise.

Human-based cognitive neuroscience and psychology, and animal model-based systems neuroscience, have complementary limitations; therefore, both contain essential pieces of the puzzle for understanding consciousness. Human participants can rapidly learn complex tasks that

isolate and control for key psychological constructs, but we lack viable methods for precisely recording and controlling neural activity in humans. At the same time, animal models - and transgenic mouse models in particular - allow for an astonishing degree of experimental precision in the recording and causal manipulation of neural activity, but are highly limited in the range and complexity of the tasks they can perform, restricting the type of psychological inferences that can be drawn. Effective progress in the neuroscience of consciousness, therefore, hinges on our ability to create empirically tractable tethers between the behaviourally precise signatures of consciousness studied in humans and the fine-grained neurobiological mechanisms studied in animal models (cf. [37, 38, 57]).

1.3 Aims and scope: quantitative models are a precondition for effective theory

The aim of this thesis is to begin to bridge this explanatory gap by constructing biophysical models, at multiple scales, out of component parts specified by systems neuroscience, that are able to perform tasks that have been studied and validated in human neuroscience and psychophysics. Importantly, I will limit my scope to the subset of questions concerned with the contents of consciousness, in particular, to the neuronal transition underlying the change from unconscious to conscious perceptual processing. The motivation behind this limit in scope is twofold. First, of the three general research programs studied in the neuroscience of consciousness, it is only possible to reliably study the contents and global state of consciousness in animal models. Animal models of conscious self-related phenomena are emerging [58], but are far from being well-developed enough to benefit from biophysical modelling. Additionally, it will likely never be possible to study the more elaborate aspects of self-related processing that require language in animal models.

Second, because of its immediate clinical relevance, the global state of consciousness (in the form of general anaesthesia and sleep) has been the target of biophysical modelling for decades [59–64]. Providing novel, experimentally useful insight is, therefore, more difficult to achieve within the confines of a single PhD. In contrast, the contents of consciousness, and in particular the subset of questions related to perceptual awareness, sit in a "Goldilocks zone" for potential theoretical progress. There is a large and rapidly growing body of experimental work in humans and animal models, and a comparable scarcity of quantitative neurobiological models. The field is, therefore, ideally situated to benefit from biophysically detailed modelling.

Importantly, instead of trying to extend or compare existing theories of consciousness, I will focus on modelling and extending the results of individual experimental paradigms. I (somewhat optimistically) suspect that the neuroscience of consciousness is at a similar stage of scientific development to physics at the beginning of the twentieth century. We have a lot of difficult-to-explain experimental results, and no (satisfactory) unifying explanatory framework. Notably, the development of quantum mechanics occurred after, and arguably because of, developments made whilst trying to model the results of individual experiments. When Planck proposed that energy changes in discrete quantised jumps, he was not trying to propose a new unified theory of matter. Rather, he was quite specifically concerned with trying to model what was, at that point, an idiosyncratic experimental result: the spectra of black-body radiation [65]. There was an enormous amount of experimental work and physical modelling between Planck's initial insight published at the turn of the 20th century [66] and the development of fully-fledged quantum mechanics in 1927 [67]. I am of the opinion, therefore, that

we are not at the appropriate stage of scientific development to build theories of consciousness. Instead, we need to have successful neuronal models of the many and varied individual phenomena we are trying to explain before we try to construct any kind of unified theory, and we are currently very far away from achieving this goal. In light of this, I focus on building models that are as close to the relevant experimental data as possible, and always propose, and (where possible) also test, additional empirical predictions.

1.4 Essential elements for a computational neurobiology of perceptual awareness

The success of this approach relies upon two key elements. First, models need to balance neurobiological detail with interpretability and give insights that transcend the scale at which the model is proposed. I will, therefore, rely heavily on tools from nonlinear dynamical systems theory, which can be applied at every scale from multicompartment models of single-neuron dynamics to rate-based neural mass and mean-field models, which describe the mass action of neuronal populations consisting of tens of thousands of neurons, all the way up to regions of tens of millions of neurons. Across these vastly disparate scales, the dynamics are described by nonlinear differential equations that, despite not having analytic solutions, can be rigorously understood in terms of the existence and stability of the attractors and equilibria governing their behaviour. The name of the game is then identifying the parameters of the system that control the existence and stability of attractors. As a parameter is smoothly varied, if attractors abruptly change their stability or come into or out of existence, the system has passed through a bifurcation [68, 69]. Bifurcations show up everywhere in nature, from simple quantum mechanical systems such as Josephson Junctions [69, 70] to large-scale models of ecosystem collapse [71]. Knowing that a system is next to a particular type of bifurcation provides us with a huge amount of information. If a system is sufficiently close to a bifurcation, and we know the type of bifurcation the system is next to, the dynamics are topologically equivalent (i.e., have qualitatively identical dynamics) to a simple algebraic system, known as a normal form, that is independent of the idiosyncratic details of the system in question [68, 69, 72]. Bifurcation theory, therefore, provides a rare window of interpretability in the face of the nervous system's vast and formidable degeneracy. As such, throughout this thesis, I go to great effort to find the perspective most amenable to the application of bifurcation theory. Sometimes we will study the behaviour of sub-compartments of single neurons, while at other times we will study the dynamics of the system as a whole. Throughout, my goal remains the same: to identify the neurobiological bifurcation parameters governing the brain's dynamics.

The second key element is the existence of psychophysical constraints. In order to be confident in generalising from animal models to humans, we need to study behaviour with specific and highly reproducible signatures so that we can be confident we are studying the same phenomenon in (typically overtrained) animals as we are in human participants. I will, therefore, limit the scope of the thesis to two key families of psychophysical paradigms in the neuroscience of consciousness: threshold detection and visual rivalry. Threshold detection is about as simple as tasks come: a simple sensory stimulus is presented to a subject, over a range of intensities, and they are then asked to indicate whether the stimulus was present or absent (n.b., the stimulus is withheld on a subset of trials to control for response bias). Response rates follow a sigmoidal function with the perceptual threshold - the stimulation intensity at which subjects report the stimulus ∼50% of the time - occurring at the inflection point of the sigmoid [73]. Sorting trials at a subject's perceptual threshold into conscious and unconscious

conditions provides a physically-matched comparison between conditions where a subject was conscious of the stimulus, and a condition where the stimulus remained unconscious, failing to pass the internal threshold for awareness [29]. Threshold detection paradigms can be used across a variety of sensory modalities, making it suitable for animal models for whom vision is not the dominant sense. The task is also simple enough that it is possible to train a variety of species to perform it. Threshold detection has, therefore, been used to study perceptual awareness in a variety of human [74] and non-human animals, including mice [43, 44] and macaques [75].

In visual rivalry, each eye is presented with physically-matched but incongruent images (e.g., orthogonal gratings), and instead of perceiving a fusion of the two images, binocular fusion is disrupted; one eye's image is suppressed from awareness whilst the other dominates the content of perception in a process known as interocular suppression [76–79]. Perception then alternates stochastically between the two percepts between periods of dominance and suppression every few seconds [80, 81]. The appeal of this phenomenon is simple: two visual stimuli enter the visual system, but only one is consciously experienced, providing a practical means to study the processes underlying visual awareness and suppression. Visual rivalry has a long experimental history in both human psychophysics [79, 82] and non-human primate electrophysiology [83–89]. Crucially, the relationship between stimulus intensity and the duration of the resulting dominance and suppression periods is highly regular and can be captured by a small set of law-like propositions (known as Levelt's laws; [90, 91]) which have proven remarkably robust, needing only minor modification and revision since they were first proposed over 60 years ago. In addition, the distribution of dominance durations of each competing percept has a characteristic right-skew that is well described by gamma and lognormal distributions [80, 81].

With the right choice of stimulus, rivalry paradigms also do not require subjects to make overt behavioural responses. Involuntary eye movements (optokinetic nystagmus) reliably track the dominant percept [84, 92, 93], making it ideal for use in animal models (e.g. [83, 85]). Indeed, Levelt's laws and the right-skewed distribution of dominance durations have both been shown to hold in human and non-human primates alike [90, 94], and while the presence of interocular competition in mouse models is still a matter of ongoing research [95], monocular rivalry paradigms that utilise competing incongruent motion stimuli have been employed in mice and display the same right-skewed distribution of perceptual dominance durations [96, 97]. Visual rivalry paradigms, therefore, offer a second paradigm beyond threshold detection that can reliably be studied in human and animal models alike, allowing us to construct explanatory bridges between detailed neuronal mechanisms and psychological phenomena. There are other paradigms that hold physical stimulation constant whilst varying the content of perception, such as visual masking [98] and illusory contours [99], that are starting to be used in animal models. However, it is still early days, and so for the sake of simplicity, I will restrict myself to threshold detection and visual rivalry as they are by far the two most well-studied and reliable paradigms that can be employed across human participants and animal models.

Here, I adopt the term "cross-level mechanistic theory", recently introduced by Wang [100], to describe the approach to modelling taken in this thesis. Despite my misgivings about the use of the word 'theory' in this context for the reasons outlined above, we are closely aligned on the use of nonlinear dynamics to bridge scales in neural systems. In addition, we share a focus on the importance of studying elemental cognitive functions that can be investigated in animal models, and on the importance of starting with careful quantitative modelling of individual experiments, rather than proposing unifying theories based upon an abstract characterisation

of a phenomenon. As such, I will put semantic quibbles aside and make heavy use of the term throughout the thesis.

1.5 Thesis structure

This thesis is divided into two parts. In part one, I build on recent work in the systems neuroscience of consciousness, beginning with a review in chapter two of the current state of the art in systems and computational neuroscience. I review evidence showing that thalamocortical loops play a demonstrable role in the global state of consciousness, and although more controversial, further propose that the current state of the evidence supports the view that thalamocortical loops also play an essential role in determining the contents of perceptual consciousness. In chapter three, I construct a spiking neural network model of perceptual awareness that incorporates essential features of thalamocortical anatomy and cellular physiology. The model reproduces, and mechanistically explains, the key *in vivo* neural and behavioural signatures of perceptual awareness in the mouse model, as well as the response to a set of causal perturbations. I then generalise the same model with identical parameters to a more complex task - visual rivalry - and find that the same thalamic-mediated mechanism of perceptual awareness determines perceptual dominance, leading to the generation of a set of novel electrophysiological predictions.

In part two, I switch focus from the thalamocortical neurobiology of perceptual awareness to the dynamical principles governing perceptual awareness and suppression in visual rivalry. As I argued above, psychophysical constraints are crucial when building complex models. Showing that a model is consistent with known psychophysical findings is essential in the process of demonstrating that a model has a sufficient degree of face validity for it to be worthwhile to generate, and potentially test, neural predictions. I, therefore, take a step back from the details of thalamocortical neurobiology to study more abstract mathematically tractable models with the aim of establishing a set of psychophysical constraints, written in terms of law-like closed-form expressions. In particular, I focus on a recently proposed visual rivalry paradigm, tracking continuous flash suppression [101], which allows researchers to measure the stimulus intensity threshold for visual awareness and suppression. The paradigm is fast, robust, and behaviourally simple, making it a promising candidate for use in animal models, and in particular non-human primates.

In chapter four, I show that under a reasonable set of assumptions, one can derive a reduced, purely cortical, neural mass model of visual rivalry from a thalamocortical neural mass model that more faithfully represents the dynamics of the circuitry we studied in part one. Having established the validity of the simplified system, in chapters five and six, I extend an existing model of visual rivalry to tracking continuous flash suppression. In chapter five, I use numerical simulations to show that the same mechanisms that govern perceptual awareness and suppression in binocular rivalry - the most common and well-understood visual rivalry paradigm - generalise to tracking continuous flash suppression. In chapter six, I then derive law-like closed-form expressions from the same model for dominance and suppression durations, and for awareness and suppression thresholds, which I leverage to propose and test a novel behavioural prediction in previously collected psychophysical data [101]. It is my hope that this will serve as a foundation for a micro-scale thalamocortical spiking model of tracking continuous flash suppression, potentially providing us with a general cellular-level explanation of the threshold for visual awareness and suppression that can be tested in animal models.

Finally, I close the thesis in chapter seven with a discussion of the overarching and recurring themes encountered throughout. We explore the question of when and why it is important to include cellular-level detail in computational models. I highlight the relationship between the models and findings explored in the thesis and existing neuroscientific theories of consciousness. I conclude with a speculative discussion of what a mature cross-level mechanistic (thalamocortical) theory of conscious contents might look like.

Chapter 1 references

[1] Anil Seth. *Being you: A new science of consciousness*. Penguin Publishing Group, 2021.

[2] Bernard J Baars. "The conscious access hypothesis: Origins and recent evidence". In: *Trends in cognitive sciences* 6.1 (2002), pp. 47–52.

[3] Graham Teasdale et al. "The Glasgow Coma Scale at 40 years: standing the test of time". In: *The Lancet Neurology* 13.8 (2014), pp. 844–854.

[4] Tim Bayne, Jakob Hohwy, and Adrian M Owen. "Are there levels of consciousness?" In: *Trends in cognitive sciences* 20.6 (2016), pp. 405–413.

[5] Shaun Gallagher. "Philosophical conceptions of the self: Implications for cognitive science". In: *Trends in cognitive sciences* 4.1 (2000), pp. 14–21.

[6] Anna Ciaunica et al. "I overthinktherefore i am not: An active inference account of altered sense of self and agency in depersonalisation disorder". In: *Consciousness and cognition* 101 (2022), p. 103320.

[7] Anil K Seth. "Interoceptive inference, emotion, and the embodied self". In: *Trends in cognitive sciences* 17.11 (2013), pp. 565–573.

[8] Anil K Seth and Manos Tsakiris. "Being a beast machine: The somatic basis of selfhood". In: *Trends in cognitive sciences* 22.11 (2018), pp. 969–981.

[9] Anil K Seth and Tim Bayne. "Theories of consciousness". In: *Nature Reviews Neuroscience* 23.7 (2022), pp. 439–452.

[10] Bernard J Baars. "Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience". In: *Progress in brain research*. Vol. 150. Elsevier, 2005, pp. 45–53.

[11] Bernard J Baars, Stan Franklin, and Thomas Z Ramsoy. "Global workspace dynamics: Cortical binding and propagation enables conscious contents". In: *Frontiers in psychology* 4 (2013), p. 200.

[12] Stanislas Dehaene, Jean-Pierre Changeux, and Lionel Naccache. "The global neuronal workspace model of conscious access: From neuronal architectures to clinical applications". In: *Characterizing consciousness: From cognition to the clinic?* Springer, 2011, pp. 55–84.

[13] George A Mashour et al. "Conscious processing and the global neuronal workspace hypothesis". In: *Neuron* 105.5 (2020), pp. 776–798.

[14] Stanislas Dehaene et al. "Conscious, preconscious, and subliminal processing: a testable taxonomy". In: *Trends in cognitive sciences* 10.5 (2006), pp. 204–211.

[15] Tim Bayne and Olivia Carter. "Dimensions of consciousness and the psychedelic state". In: *Neuroscience of Consciousness* 2018.1 (2018), niy008.

[16] Richard Brown, Hakwan Lau, and Joseph E LeDoux. "The misunderstood higher-order approach to consciousness". In: *PsyArXiv* (2019).

[17] Stephen M Fleming. "Awareness as inference in a higher-order state space". In: *Neuroscience of consciousness* 2020.1 (2020), niz020.

[18] Hakwan Lau and David Rosenthal. "Empirical support for higher-order theories of conscious awareness". In: *Trends in cognitive sciences* 15.8 (2011), pp. 365–373.

[19] Larissa Albantakis et al. "Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms". In: *PLoS computational biology* 19.10 (2023), e1011465.

[20] Masafumi Oizumi, Larissa Albantakis, and Giulio Tononi. "From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0". In: *PLoS computational biology* 10.5 (2014), e1003588.
[21] Giulio Tononi et al. "Integrated information theory: from consciousness to its physical substrate". In: *Nature Reviews Neuroscience* 17.7 (2016), pp. 450–461.
[22] J Kevin O'Regan and Alva Noë. "A sensorimotor account of vision and visual consciousness". In: *Behavioral and brain sciences* 24.5 (2001), pp. 939–973.
[23] Anil K Seth. "A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia". In: *Cognitive neuroscience* 5.2 (2014), pp. 97–118.
[24] Thomas Metzinger. *Being no one: The self-model theory of subjectivity*. MIT press, 2004.
[25] David Rudrauf et al. "A mathematical model of embodied consciousness". In: *Journal of theoretical biology* 428 (2017), pp. 106–131.
[26] I Yaron et al. "The ConTraSt database for analysing and comparing empirical studies of consciousness theories". In: *Nature human behaviour* 6.4 (2022), pp. 1–7.
[27] Astrid A Prinz, Dirk Bucher, and Eve Marder. "Similar network activity from disparate circuit parameters". In: *Nature neuroscience* 7.12 (2004), pp. 1345–1352.
[28] John W Krakauer et al. "Neuroscience needs behavior: correcting a reductionist bias". In: *Neuron* 93.3 (2017), pp. 480–490.
[29] Morten Overgaard. *Behavioural methods in consciousness research*. Oxford University Press, 2015.
[30] Stanislas Dehaene. *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin Publishing Group, 2014.
[31] Victor AF Lamme. "Towards a true neural stance on consciousness". In: *Trends in cognitive sciences* 10.11 (2006), pp. 494–501.
[32] INTREPID CONSORTIUM. *OSF Registries | Accelerating research on consciousness: An adversarial collaboration to test contrasting predictions of the Integrated Information Theory and Predictive Processing accounts of consciousness*. https://osf.io/35rhx. 2021.
[33] Nicholas A Steinmetz et al. "Challenges and opportunities for large-scale electrophysiology with Neuropixels probes". In: *Current opinion in neurobiology* 50 (2018), pp. 92–100.
[34] Christine Grienberger et al. "Two-photon calcium imaging of neuronal activity". In: *Nature Reviews Methods Primers* 2.1 (2022), p. 67.
[35] Lief Fenno, Ofer Yizhar, and Karl Deisseroth. "The development and application of optogenetics". In: *Annual review of neuroscience* 34 (2011), pp. 389–412.
[36] Daniel J Urban and Bryan L Roth. "DREADDs (Designer Receptors Exclusively Activated by Designer Drugs): chemogenetic tools with therapeutic utility". In: *Annual review of pharmacology and toxicology* 55 (2015), pp. 399–417.
[37] Biyu J He. "Next frontiers in consciousness research". In: *Neuron* 111.20 (2023), pp. 3150–3153.
[38] Johan F Storm et al. "Consciousness regained: disentangling mechanisms, brain systems, and behavioral responses". In: *Journal of Neuroscience* 37.45 (2017), pp. 10882–10893.
[39] Jaan Aru et al. "Coupling the state and contents of consciousness". In: *Frontiers in systems neuroscience* 13 (2019), p. 43.

[40] Jaan Aru et al. “Apical drivea cellular mechanism of dreaming?” In: *Neuroscience & Biobehavioral Reviews* 119 (2020), pp. 440–455.

[41] Jaan Aru, Mototaka Suzuki, and Matthew E Larkum. “Cellular mechanisms of conscious processing”. In: *Trends in cognitive sciences* 24.10 (2020), pp. 814–825.

[42] Mototaka Suzuki and Matthew E Larkum. “General anesthesia decouples cortical pyramidal neurons”. In: *Cell* 180.4 (2020), 666–676.e13.

[43] Naoya Takahashi et al. “Active cortical dendrites modulate perception”. In: *Science* 354.6319 (2016), pp. 1587–1590.

[44] Naoya Takahashi et al. “Active dendritic currents gate descending cortical outputs in perception”. In: *Nature neuroscience* 23.10 (2020), pp. 1277–1285.

[45] Francisco Clascá et al. “Anatomy and development of multispecific thalamocortical axons”. In: *Axons and brain architecture*. Elsevier, 2016, pp. 69–92.

[46] Kenneth D Harris and Gordon MG Shepherd. “The neocortical circuit: themes and variations”. In: *Nature neuroscience* 18.2 (2015), pp. 170–181.

[47] Edward G Jones. “The thalamic matrix and thalamocortical synchrony”. In: *Trends in neurosciences* 24.10 (2001), pp. 595–601.

[48] Gordon MG Shepherd and Naoki Yamawaki. “Untangling the cortico-thalamo-cortical loop: cellular pieces of a knotty circuit puzzle”. In: *Nature Reviews Neuroscience* 22.7 (2021), pp. 389–406.

[49] James M Shine. “Adaptively navigating affordance landscapes: How interactions between the superior colliculus and thalamus coordinate complex, adaptive behaviour”. In: *Neuroscience & Biobehavioral Reviews* 143 (2022), p. 104921.

[50] Matthew E Larkum. “Are dendrites conceptually useful?” In: *Neuroscience* 489 (2022), pp. 4–14.

[51] Matthew E Larkum. “Top-down dendritic input increases the gain of layer 5 pyramidal neurons”. In: *Cerebral cortex* 14.10 (2004), pp. 1059–1070.

[52] Matthew E Larkum et al. “Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle”. In: *Science* 325.5941 (2009), pp. 756–760.

[53] Matthew E Larkum et al. “The guide to dendritic spikes of the mammalian cortex in vitro and in vivo”. In: *Neuroscience* 489 (2022), pp. 15–33.

[54] Benjamin R Munn et al. “A thalamocortical substrate for integrated information via critical synchronous bursting”. In: *Proceedings of the National Academy of Sciences* 120.46 (2023), e2308670120.

[55] Eero J Müller et al. “The non-specific matrix thalamus facilitates the cortical information processing modes relevant for conscious awareness”. In: *Cell reports* 42.8 (2023).

[56] Michelle J Redinbaugh et al. “Thalamus modulates consciousness via layer-specific control of cortex”. In: *Neuron* 106.1 (2020), 66–75.e12.

[57] Ege Kingir and Melanie Wilke. *Seeing Beyond the Human: Challenges and Advances in Animal Studies of Visual Consciousness*. 2025.

[58] M Wada et al. “The rubber tail illusion as evidence of body ownership in mice”. In: *The Journal of Neuroscience* 36.43 (2016), pp. 11133–11137.

[59] Jonathan S Goldman et al. “Bridging single neuron dynamics to global brain states”. In: *Frontiers in systems neuroscience* 13 (2019).

[60] Jonathan S Goldman et al. “A comprehensive neural simulation of slow-wave sleep and highly responsive wakefulness dynamics”. In: *Frontiers in Computational Neuroscience* 16 (2023).

[61] ML Steyn-Ross. "Toward a theory of the general-anesthetic-induced phase transition of the cerebral cortex. I. A thermodynamics analogy". In: *Physical Review E* 64.1 (2001), p. 011917.
[62] ML Steyn-Ross et al. "Theoretical electroencephalogram stationary spectrum for a white-noise-driven cortex: Evidence for a general anesthetic-induced phase transition". In: *Physical Review E* 60.6 (1999), p. 7299.
[63] ML Steyn-Ross, DA Steyn-Ross, and JW Sleigh. "Modelling general anaesthesia as a first-order phase transition in the cortex". In: *Progress in biophysics and molecular biology* 85.2 (2004), pp. 369–385.
[64] D-P Yang et al. "Wake-sleep transition as a noisy bifurcation". In: *Physical Review E* 94.2 (2016), p. 022412.
[65] James V Stone. *The quantum menagerie: A tutorial introduction to the mathematics of quantum mechanics*. Sebtel Press, 2020.
[66] Max Planck. "On an improvement of Wien's equation for the spectrum". In: *Verh. Deut. Phys. Ges* 2 (1900), pp. 202–204.
[67] Sean Carroll. *Something deeply hidden: Quantum worlds and the emergence of spacetime*. Penguin Books, 2019.
[68] Eugene M Izhikevich. *Dynamical systems in neuroscience: The geometry of excitability and bursting*. The MIT Press, 2006.
[69] Steven Strogatz. *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Second. CRC press, 2018.
[70] Brian D Josephson. "Possible new effects in superconductive tunnelling". In: *Physics letters* 1.7 (1962), pp. 251–253.
[71] Marten Scheffer and Stephen R Carpenter. "Catastrophic regime shifts in ecosystems: linking theory to observation". In: *Trends in ecology & evolution* 18.12 (2003), pp. 648–656.
[72] John Guckenheimer and Philip Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Springer Science & Business Media, 2013.
[73] Donald McNicol. *A primer of signal detection theory*. Psychology Press, 2005.
[74] Zhiyue Fang et al. "Human high-order thalamic nuclei gate conscious perception through the thalamofrontal loop". In: *Science* 388.6742 (2025), eadr3675.
[75] B van Vugt et al. "The threshold for conscious report: Signal loss and response bias in visual and frontal cortex". In: *Science* 360.6388 (2018), pp. 537–542.
[76] David Alais and Randolph Blake. *Binocular rivalry*. MIT press, 2005.
[77] Randolph Blake. "A primer on binocular rivalry, including current controversies". In: *Brain and Mind* 2.1 (2001), pp. 5–38.
[78] Randolph Blake and K Boothroyd. "The precedence of binocular fusion over binocular rivalry". In: *Perception & Psychophysics* 37.2 (1985), pp. 114–124.
[79] Randolph Blake and Nikos K Logothetis. "Visual competition". In: *Nature reviews neuroscience* 3.1 (2002), pp. 13–21.
[80] Robert Fox and J Herrmann. "Stochastic properties of binocular rivalry alternations". In: *Perception & Psychophysics* 2.9 (1967), pp. 432–436.
[81] Willem JM Levelt. "Note on the distribution of dominance times in binocular rivalry". In: *British Journal of Psychology* 58.1-2 (1967), pp. 143–145.
[82] David Alais. "Binocular rivalry: Competition and inhibition in visual perception". In: *Wiley Interdisciplinary Reviews: Cognitive Science* 3.1 (2012), pp. 87–103.
[83] Abhilash Dwarakanath et al. "Prefrontal state fluctuations control access to consciousness". In: *bioRxiv* (2020).

[84] Jan K Hesse and Doris Y Tsao. "A new no-report paradigm reveals that face cells encode both consciously perceived and suppressed stimuli". In: *eLife* 9 (2020), e58360.
[85] Vishal Kapoor et al. "Decoding the contents of consciousness from prefrontal ensembles". In: *bioRxiv* (2020).
[86] David A Leopold and Nikos K Logothetis. "Activity changes in early visual cortex reflect monkeys percepts during binocular rivalry". In: *Nature* 379.6565 (1996), pp. 549–553.
[87] Nikos K Logothetis. "Single units and conscious vision". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 353.1377 (1998), pp. 1801–1818.
[88] Theofanis I Panagiotaropoulos et al. "Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex". In: *Neuron* 74.5 (2012), pp. 924–935.
[89] Melanie Wilke, K-M Mueller, and David A Leopold. "Neural activity in the visual thalamus reflects perceptual suppression". In: *Proceedings of the National Academy of Sciences* 106.23 (2009), pp. 9465–9470.
[90] Jan W Brascamp, P Christiaan Klink, and Willem JM Levelt. "The laws of binocular rivalry: 50 years of Levelts propositions". In: *Vision research* 109 (2015), pp. 20–37.
[91] Willem JM Levelt. *On binocular rivalry*. Tech. rep. Institute for Perception, 1965.
[92] Stefan Frässle et al. "Binocular rivalry: frontal activity relates to introspection and action but not to perception". In: *Journal of Neuroscience* 34.5 (2014), pp. 1738–1747.
[93] Marnix Naber, Stefan Frässle, and Wolfgang Einhäuser. "Perceptual rivalry: reflexes reveal the gradual nature of visual awareness". In: *PLoS one* 6.6 (2011), e20910.
[94] Nikos K Logothetis, David A Leopold, and DL Sheinberg. "What is rivalling during binocular rivalry?" In: *Nature* 380.6575 (1996), pp. 621–624.
[95] M Timplalexi, WM Connelly, and A Ranson. "Feedforward and feedback population dynamics during binocular conflict in mouse visual cortex". In: *bioRxiv* (2025).
[96] Daria Bogatova, Stelios M Smirnakis, and Galina Palagina. "Tug-of-peace: Visual Rivalry and Atypical Visual Motion Processing in MECP2 duplication Syndrome of Autism". In: *bioRxiv* (2022).
[97] Galina Palagina, J F Meyer, and Stelios M Smirnakis. "Complex visual motion representation in mouse area V1". In: *The Journal of Neuroscience* 37.1 (2017), pp. 164–183.
[98] S David Gale et al. "Backward masking in mice requires visual cortex". In: *Nature Neuroscience* 27.1 (2024), pp. 129–136.
[99] H Shin et al. "Recurrent pattern completion drives the neocortical representation of sensory inference". In: *Nature Neuroscience* (2025), pp. 1–11.
[100] Xiao-Jing Wang. *Theoretical neuroscience: Understanding cognition*. CRC press, 2025.
[101] David Alais et al. "A new CFS tracking paradigm reveals uniform suppression depth regardless of target complexity or salience". In: *eLife* 12 (2024), RP91019.

Part 1

The Thalamocortical Basis of Perceptual Awareness

Chapter 2

Thalamocortical Control of the State and Contents of Consciousness

> The task of philosophy abstractly formulated is to understand how things, in the broadest possible sense of the term, hang together, in the broadest possible sense of the term.
>
> — *Wilfred Sellars*

In his famous 1962 paper "Philosophy and the Scientific Image of Man", Wilfred Sellars introduced a distinction between what he called the manifest and scientific images of the world [1]. To paraphrase Daniel Dennett [2], the manifest image is the world we live in; it is the world of tables and chairs, teacups and Tuesdays, love and despair. The scientific image is the world described by our best scientific theories; it is the world of atoms and molecules, electrons and neutrons, neurons and action potentials. The task of philosophy, according to Sellars, is to understand how these two worlds hang together.

I think there is a sense in which explaining this distinction is also the primary goal of the neuroscience of consciousness. Unconscious processes are by their very nature something that we must infer from a third-person (scientific) perspective; they are not a part of our experience. They are, in this sense, a part of the scientific image. Conscious processes are, by the same token, a part of the manifest image. What we want is a neuronal bridge between the scientific and manifest images.

In this chapter I, along with my coauthors Michelle Redinbaugh, Mac Shine, and Yuri Saalman, review the state of the art in the systems, cellular, and computational neuroscience of consciousness, to argue that, according to the best evidence currently available the thalamus, and thalamocortical loops in general, control both the state and contents of consciousness. That is, they control the transition between both local and global states of consciousness and unconsciousness; between the manifest and scientific image. We start by reviewing the role of thalamocortical loops in controlling conscious state, which is definitive and should be relatively uncontroversial, and then extend the argument to the gating of perceptual contents

into consciousness, a more controversial (but arguably no less justified) position. In the conclusion to the thesis, we will examine an extension of the arguments given in this chapter to the properties of individual conscious contents themselves.

2.1 Chapter 2 introduction

Different functional modes of the brain vary profoundly in terms of whole-brain dynamics, their associated computational capacities, and the nature of the corresponding conscious experiences. These different modes of conscious processing can be framed as varying across at least two axes [3–5]: *conscious state* - an organisms level of arousal and associated global changes in behavioral state, such as wakefulness, sleep, and anesthesia; and *conscious content* - the relatively local variations in the informational composition of an individual conscious experience that occur against the backdrop of a particular global state of consciousness, including moment-to-moment changes in both the content of, and threshold for, awareness.

Progress has been made regarding the neural mechanisms that together support both conscious state and contents - from classical animal studies of state, to more recent theoretically-driven human neuroimaging studies of content. However, the lack of consensus regarding detailed neurobiological mechanisms means that it is exceedingly difficult to differentiate among competing theories of consciousness [6], with identical results often interpreted as supporting apparently incompatible theories [7]. Here, we review a growing body of work in systems, cellular, and theoretical neuroscience to suggest a neural foundation of cellular and systems-level processes that support conscious experience [8–16]. We survey a rich literature outlining the contribution of thalamocortical loops to the *state* of consciousness, such as the distinction between sleep, wake, and anesthesia, drawing empirical evidence from a wide variety of model organisms and computational approaches. We next appraise more recent empirical and computational evidence for a thalamocortical contribution to the *content* of consciousness, specifically in gating whether perceptual contents become conscious, and offer a set of provisional conclusions designed to provoke further empirical investigation. By integrating neurobiological insights across scales and model organisms, we argue that dynamic interactions between specific microcircuits connecting the thalamus and cerebral cortex form the basis of the micro-, meso- and macro-scopic neural dynamics that support both the state and content of consciousness.

2.2 The anatomical basis of thalamocortical connectivity

The thalamus is a highly-conserved subcortical structure that is robustly and precisely interconnected with the rest of the brain at the micro-scale (Figure 7.1A), meso-scale (Figure 7.1B/C) and macro-scale (Figure 7.1D).

At the microscopic level, the thalamus is intimately connected with the cerebral cortex. Cortico-thalamic projections originate from the deep, infragranular layers (thick-tufted L5 extratelencephalic ($L5_{ET}$) and L6 cells) while reciprocal thalamo-cortical projections predominantly terminate in either the middle (granular L4 and deep L3) or superficial, supragranular (L1) layers, with relatively sparse (though potentially impactful [17, 18]) projections to the deeper layers (Figure 7.1A). Thalamic neurons projecting to the middle cortical layers (known

as core neurons) typically contain the calcium binding protein parvalbumin [19–21]. In contrast, thalamic neurons projecting to the superficial or deep cortical layers (known as matrix neurons) contain the calcium binding protein calbindin [19–21]. Core neurons tend to have strong effects on their cortical targets - i.e., cause large excitatory postsynaptic potentials (EPSPs) via ionotropic glutamate receptors - and thus are considered drivers. Matrix neurons tend to have weaker, graded effects - i.e., smaller EPSPs via both metabotropic and ionotropic glutamate receptors - and are thus considered modulators [22, 23]. Spiking activity from the sensory organs is transmitted to the thalamus (mainly via first-order thalamic nuclei, such as the lateral geniculate nucleus or medial geniculate body), which in turn sends dense axonal projections to primary sensory cortices. Information transmission between cortical neurons occurs via direct cortico-cortical connections, and via higher-order thalamic nuclei (e.g., the pulvinar and intralaminar nuclei) that interconnect wide-spread cortical areas via robust cortico-thalamo-cortical pathways [24].

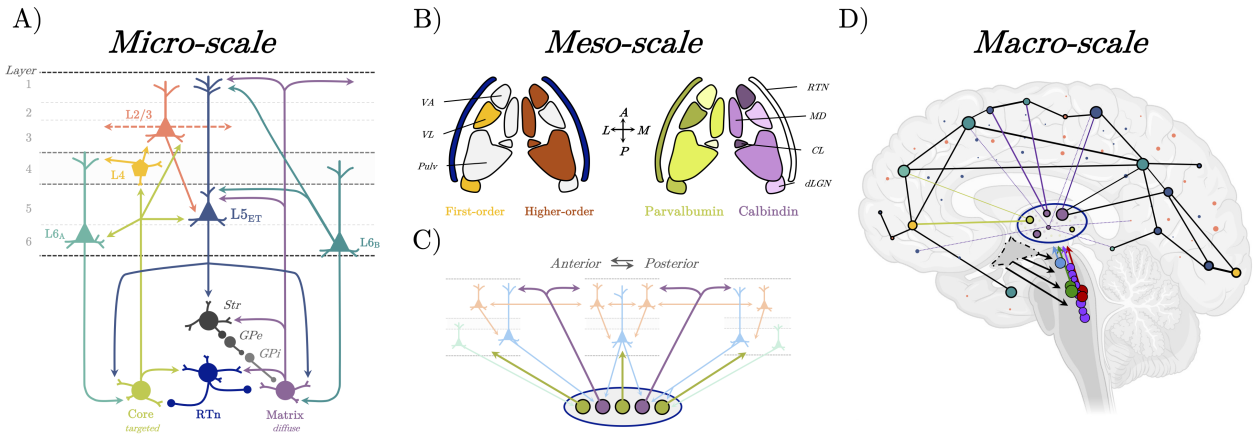

Figure 2.1: **A)** At the micro-scale, distinct excitatory cell types in the thalamus project to distinct layers of the cerebral cortex. **B)** The different nuclei of the thalamus characterized according to First-Order/Higher-Order status and expression of parvalbumin and calbindin. **C)** At the meso-scale, the thalamus controls local and long-range excitability of the cerebral cortex. **D)** At the macro-scale, the thalamus is deeply interconnected with the entire cortical mantle, forming a distributed network with convergent and divergent architectures.

At the mesoscopic scale, the thalamus has been divided into at least 30 nuclei, based on connectivity as well as cytoarchitectonic, neurochemical, and functional criteria [25]. The distribution of core and matrix neurons differs across thalamic nuclei (Figure 7.1B): core neurons predominate in first-order thalamic nuclei that receive driver-inputs from the sensory periphery, whereas matrix neurons predominate in the higher-order intralaminar thalamic nuclei (except in the centromedian nucleus [CM] of the posterior intralaminar thalamus, which is parvalbumin-rich [26, 27]). Most thalamic nuclei contain a blend of both cell types, with different proportions across nuclei (Figure 7.1B, right) [27, 28]. Fascinatingly, there are also thalamic neurons that exhibit both core- and matrix-like properties [21]. This includes intralaminar nuclei, many of which have broad projections to the cerebral cortex (similar to those seen in matrix neurons), but with stronger projections to the striatum [29–31]. Through the unique projections of each type of thalamic neuron, the thalamus can simultaneously influence both local and distant patterns of neuronal spiking activity (Figure 7.1C). The excitatory projection neurons of the thalamus are also embedded within a dense inhibitory network of cells that both surround and intersperse the excitatory cells. Specifically, the thalamus is bordered by the shield-like reticular nucleus, which is composed entirely of inhibitory GABAergic cells that are excited by both excitatory cortico-thalamic and thalamo-cortical projections,

and then release (inhibitory) GABA onto a broader population of thalamic cells than the population from which they were excited [32–34]. This organization is thought to dampen ongoing thalamic activity and may also play a role in more selective features of attention and consciousness [10, 34, 35].

At the macroscale, the thalamus acts as a hub connecting the ascending sensory pathways and arousal system from the brainstem with the massive projections from the cerebellum, basal ganglia and colliculi, as well as the vast reciprocal connections with the entire cerebral cortex [10, 14, 23, 36–38]. Generally speaking, directly-connected cortical areas are mirrored (with some blurring) by cortico-thalamo-cortical projections [39, 40]. In addition to acting as a key topological hub in the brain, a key feature of the cortico-thalamic projections is their convergence (Figure 7.1C/D), which forces high-dimensional activity patterns in the cerebral cortex - useful for distinguishing between similar situations and supporting a wide range of behaviors - to be compressed, giving rise to lower-dimensional patterns in the thalamus that coincide with simpler, latent variables [41, 42]. This lower-dimensional information can then be shared via divergent outputs from the thalamus to the cerebral cortex, ensuring efficient, robust information processing across the brain [10, 43]. With this key feature in mind, it is clear that the macroscopic location of the thalamus between the cortex, subcortex and ascending arousal system places the structure at a crucial nexus for understanding the neural mechanisms of conscious state and content.

2.3 Thalamic contributions to conscious state

The profound transition in conscious experience from sleep to wake reflects a corresponding change in whole-brain dynamics so pronounced it is electrically detectable from the scalp [44]. In this transition to the waking state, a regime of low-frequency, synchronized activity indicative of sleep shifts into a high-frequency, desynchronized (i.e., asynchronous, irregular) mode reflective of a high-conductance state [45]. This transition is thought to emerge from a switch-like mechanism that originates in the lateral hypothalamus [46], wherein neurochemicals, such as orexin, are released, recruiting brainstem and forebrain regions within the ascending arousal system [14, 47, 48]. Widespread projections from these areas then release neuromodulatory chemicals, such as norepinephrine, acetylcholine, serotonin, and dopamine, which globally increase neural excitability [49], motivating the transition from sleep into wakefulness [46]. The thalamus is a particularly important downstream target of the ascending arousal system, and its specific interactions with cortical circuits plays a crucial role in restoring conscious functionality to the whole brain dynamical regime.

2.3.1 Thalamocortical microcircuit contributions to conscious state

A key mechanism through which arousal based neuromodulatory inputs exert their effect is via T-type calcium channels and non-specific I_h channels in the thalamus. In the waking state, T-type channels are inactive and the cells operate in a tonic mode, firing regular action potentials at variable rhythms. At moderate levels of hyperpolarization, T-type calcium currents are open, causing an inhibition-induced burst that operates in the typical spindle frequency. Further hyperpolarization then opens non-specific I_h channels, which paradoxically depolarizes the cell, reducing the rate of inhibition-induced slow wave bursts to a delta frequency. As the cells cycle rhythmically through these up and down states, they set the pace of cortical rhythms [50–55]: first the sleep spindles that are emblematic of early stages of NREM, then

of slow wave activity (SWA) that characterizes later stages of sleep. *In vivo* thalamic manipulation experiments have further reinforced the role of the thalamus in electrophysiological signatures of brain states: thalamic inhibition in freely moving rats has been shown to alter both sleep spindles and slow cortical waves, while optogenetic activation entrains slow-wave frequencies within a narrow band (0.75-1.5Hz) [56]. Importantly, this pacemaker mechanism for cortical SWA in NREM is also characteristic of SWA in anesthetic-induced unconsciousness [57].

Why should the thalamus play such a crucial role in shaping the expression of overall brain states? The weight of evidence suggests that (neuromodulatory-controlled) thalamic bursts influence the onset and timing of cortical up-states [51, 57]. More broadly, thalamus may influence conscious state by tuning integrative properties of the cerebral cortex, such as oscillatory synchrony, neuronal resonance (responsiveness to inputs at specific frequencies), and functional connectivity [14, 58–60] (for general reviews of control of oscillatory synchrony see [61, 62]). Signals that arrive in an area during a depolarized up-state more readily translate into spiking activity, while signals that arrive during down-states may fail to depolarize cells, leading to signal loss. This expected dysfunction in communication is readily observed in perturbation experiments during less conscious states. When stimulations such as transcranial magnetic stimulation are delivered during wakefulness and rapid eye movement sleep, they produce a complex EEG response that resonates across regions, while pulses delivered during NREM sleep, anesthesia, and coma yield simple EEG responses that rapidly die out [63–66]. Similar experiments in rodents hint at a thalamocortical basis of the phenomenon, as electrical or optogenetic stimulation of the deep cortical layers (which project to thalamus; Figure 7.1), trigger complex responses in conscious states not triggered by superficial layers [67, 68]. Further evidence suggests that response complexity depends on thalamic responses to the deep-layer perturbation: in wake or REM, brief bursts in thalamus give way to rebound excitation, while in NREM or anesthesia, longer periods of thalamic bursting trigger lingering cortical down-states [68].

Neural changes associated with conscious state transitions depend on intimate connections between the cerebral cortex and matrix thalamus at the microcircuit level (Figure 2.2A). Deep layers of cortex, particularly $L5_B$ and $L6_B$, project to higher-order thalamic nuclei rich with matrix cells. These matrix cells in turn project diffusely back up to the cerebral cortex, wherein they target $L5_{ET}$ neurons both at their apical dendrites, located in L1, and proximal dendrites, located in $L5_A$ [19, 21, 69]. These $L5_{ET}$ neurons participate in cortical processing in intracolumnar, lateral, feedforward, and feedback pathways, and are thus suspected by many to play a key role in consciousness [15, 70–78]. Importantly, the matrix thalamus is ideally placed to modulate their activity (Figure 2.2A). Optogenetic stimulation experiments *in vivo* demonstrate that $L5_{ET}$'s, more-so than L2/3 pyramidal neurons, generate cortical slow-waves (1Hz) [79, 80]; for reviews, see [51, 81]. $L5_{ET}$'s also uniquely synchronize in up-down states across cortex under general anesthesia relative to neurons in other cortical layers [82]. Empirical evidence suggests thalamic bursting drives the rhythmic properties of these cells [51, 67, 83], and neuromodulatory input from the thalamus may more generally control their responsiveness to other inputs. Indeed, a recent study pairing optogenetic stimulation of dendrites in L1 with patch clamp recordings of the apical dendrites and soma of the same neurons showed that dendrites and soma are largely synchronized in response to dendritic activation during wakefulness, but decoupled under multiple types of anesthesia (Figure 2.2B) [84]. This effect could be replicated by the application of mGluR antagonists (intended to block glutamatergic thalamic input to the coupling zone that allows activity to propagate between apical and

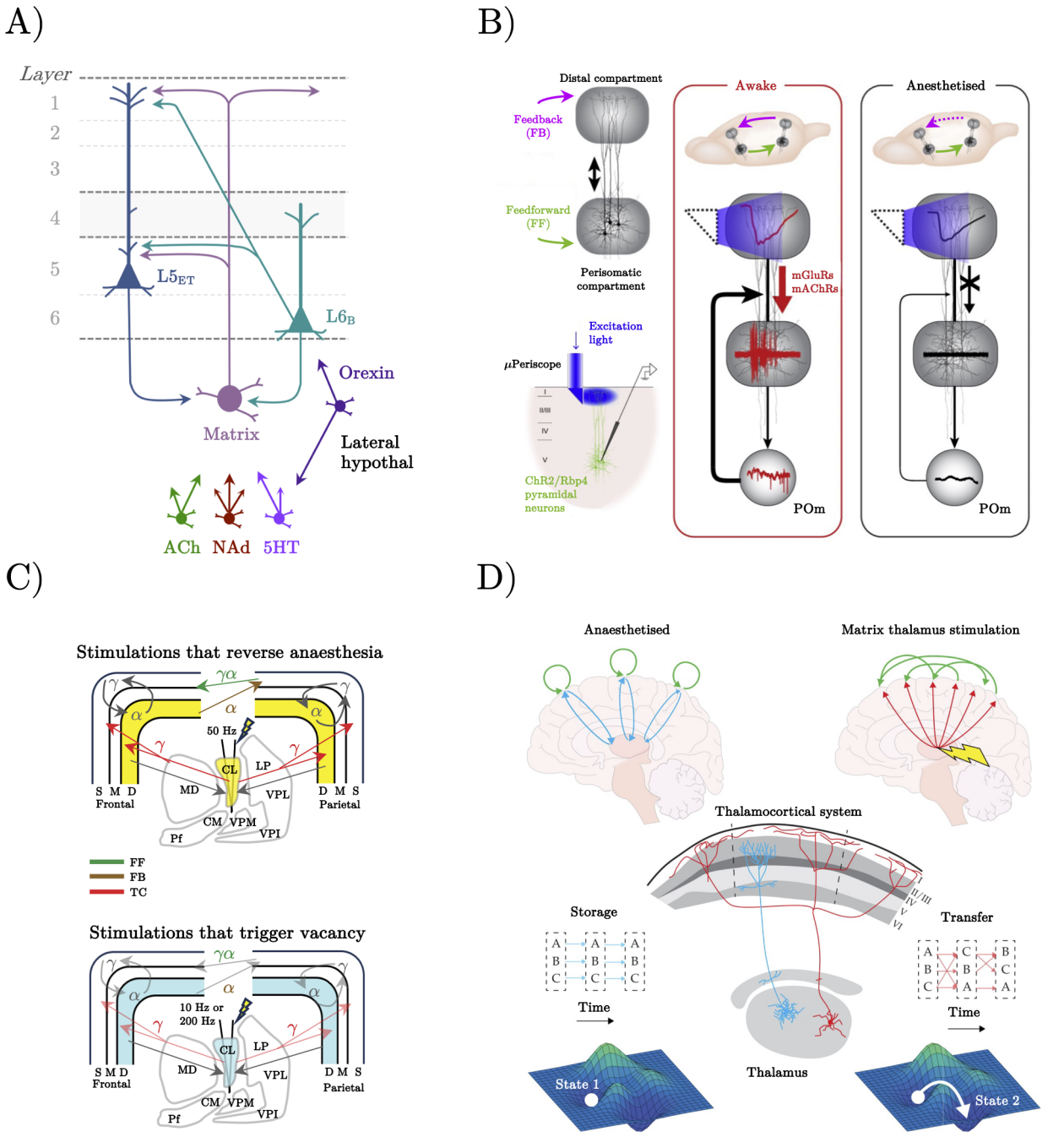

Figure 2.2: **A)** Thalamocortical circuits implicated in conscious state. **B)** During wakefulness, matrix thalamic inputs excite $L5_{ET}$ neurons. Under general anesthesia, this coupling is lost. **C)** Minimally sufficient mechanism for CL thalamic DBS to restore or disrupt consciousness. **D)** A large-scale corticothalamic neural mass model replicating recovery of consciousness with thalamic stimulation.

somatic compartments), as well as direct inactivation of the thalamus with muscimol [84]. When decoupled, $L5_{ET}$'s are unable to combine activity from feedback projections terminating in superficial layers with activity from feed forward projections terminating in deep layers [73, 75], or activity even within the cortical column. Thus, thalamic dynamics can directly contribute to a breakdown in feedforward and feedback information flow associated with a drop in conscious state [15].

Other deep layer neurons that interact with thalamus and $L5_{ET}$ neurons to influence conscious states are $L6_B$ pyramidal neurons, which provide modulatory input to the matrix-rich higher-order thalamic nuclei [17, 24, 36, 85], and also send axonal projections to L1, wherein they contact the apical dendrites of $L5_{ET}$'s (Figure 2.2A). Activation of these L1 efferents generates calcium spikes in the apical dendrites of $L5_{ET}$'s, an effect which is eliminated by the application of an NMDA-R antagonist [17]. This pathway has unique ties to conscious state and arousal because, while all cortical neurons express receptors for different classes of neuromodulatory inputs [14], $L6_B$ pyramidal neurons are the only cortical cells sensitive to orexin [17].

2.3.2 Thalamocortical mesocircuit contributions to conscious state

The breadth of thalamocortical projections expands this microcircuit motif into the mesoscale (Figure 7.1C). Stemming from a variety of higher-order thalamic nuclei, matrix efferents span the cerebral cortex and target many circuits associated with specific higher-order functions [36, 86]. During wakefulness, tonic spiking in the thalamus tunes the responsiveness of networked cortical areas to optimize efficient information transmission [59, 87], and in doing so, contributes to a responsive, functionally coherent conscious state [58]. In less conscious states, bursting thalamic activity instead sets the pace of cortical slow-waves in a way that contributes to the areas functional isolation from the rest of the network. During NREM sleep, it has been shown that distinct brain areas oscillate at different frequencies and relative phases [88]. By desynchronizing up-states between brain areas, matrix thalamic projections are well-placed to ensure the functional isolation required for modular processing [14], rendering cortical communication sparse, inconsistent, and dissociated from normal sensory pathways. As the network function shifts instead towards mechanisms of synaptic homeostasis and memory consolidation [51, 89], the resulting drop in integrated network topology and coherent transmission of content would contribute to a loss of consciousness [90, 91].

Important causal evidence for the role of thalamus in the generation and maintenance of conscious states comes from clinical observations related to disorders of consciousness, especially coma and absence epilepsy. Broadly, coma is associated with lesions that affect the brainstem, however bilateral thalamic lesions [92] and lesions that decouple the thalamus from the reticular activating system [93, 94] can also cause coma, and recovery from coma is associated with improved function in thalamocortical, thalamo-striatal, and fronto-parietal circuits [44, 95–99]. A number of studies have now demonstrated that thalamic deep brain stimulation (DBS), specifically of the central thalamus, i.e., intralaminar nuclei, leads to improvements for patients in coma [100–103] (for review see [104]). Interestingly, most studies target stimulation to the central lateral (CL) or centromedian (CM) nuclei, and many highlight beneficial clinical effects with moderate stimulation frequency (40-100Hz) [104]. This finding is particularly noteworthy because parallel clinical literature suggests low frequency ($\leq$10Hz) intralaminar stimulation instead triggers seizures characteristic of absence epilepsy [105, 106]. Ictal episodes of absence seizures are associated with loss of consciousness (individuals are often caught staring off into space), but with no discernible reduction in overall arousal. Thus, the intralaminar nuclei appear to exert bidirectional control over conscious state.

Mechanistic insights into the role of the thalamus in controlling conscious state have come from three recent studies that have used central thalamic DBS to reverse the effects of general anesthesia in macaque monkeys [107–109]. Despite methodological idiosyncrasies (for example, targeting CM vs. CL), in all cases animals were demonstrably roused from anesthesia

(i.e., they began to make voluntary movements and respond to external stimuli) following central thalamic stimulation, despite constant administration of anesthetic agents. One experiment also demonstrated restoration of neural responses to a local-global auditory oddball paradigm, implying the reinstatement of conscious processing [109]. Indeed, central thalamic DBS reversed many signatures of loss of consciousness and was shown to restore cortical interactions and functional connectivity in ways more typical of wakefulness [107–109]. Despite the fact that animals were roused by stimulation, CL DBS did not always eliminate SWA in the cerebral cortex, nor the dominance of slow-wave coherence in intra-cortical communication pathways. Specifically, 50Hz DBS targeted to CL - likely increasing local CL activity [110] - reinstated firing in the deep layers of cortex, as well as high-frequency intracolumnar, feedforward, and feedback coherence (Figure 2.2C) [107]. This finding suggests that the function of the thalamus goes beyond providing the cortex with excitatory tone. Rather, it is crucial for driving distinct functional modes in the cortex, even in spite of the typically-overwhelming inhibitory influence of anesthetics. In contrast, both lower (10Hz) and much higher frequency (200Hz) stimulation of CL - likely leading to reduced CL activity [111] - has been shown to produce absence-like events in wakeful macaques [112]. Without any noticeable drop in arousal, animals stared vacantly into space, occasionally making repetitive movements with the mouth (similar to absence seizure pathology). These vacant events were accompanied by altered intracolumnar dynamics, and a broadband breakdown in feedforward functional connectivity (Figure 2.2C) [112]. Indeed, many of the same complex network dynamics restored by 50Hz CL DBS under anesthesia were perturbed when DBS frequencies were more extreme [11, 107, 112] - i.e., more dissimilar from natural wakeful dynamics of the nucleus [54, 107, 112, 113]. These results directly suggest that abnormal thalamic activity can disrupt normal cortical function, even without a clear drop in arousal.

While the matrix thalamus clearly contributes to cortical functional modes, the intralaminar thalamus is particularly involved in the control of conscious states. However, substantial differences in the anatomical properties of different intralaminar nuclei (CL and CM) make the mechanism unclear. CL neurons are part of the anterior intralaminar group, stain strongly for calbindin [19] (signifying matrix-style projections), and target fronto-parietal circuits as well as the basal ganglia [36, 114, 115]. In contrast, CM is part of the posterior intralaminar group and predominantly stains for parvalbumin [19]. While this pattern is usually associated with core thalamic projections, most of CMs efferent projections target the basal ganglia [116] (which is unlike other core nuclei), with the few remaining cortical projections targeting motor areas and appearing relatively matrix-like [117–119]. If both nuclei influence consciousness by a shared mechanism, it likely involves restoration of fronto-parietal and basal ganglia circuits. This is consistent with the mesocircuit hypothesis that disruptions in intralaminar thalamus contribute to reduced excitation of the basal ganglia, triggering unconstrained tonic inhibition throughout the thalamus and substantial thalamocortical dysfunction [120]. If this proves true, CM may dominate the effect, as it has stronger basal ganglia connectivity. CM has also been shown to counter prepotent responses, via the thalamo-striatal pathway, enabling behavioral flexibility [121]. Similarly, CM stimulation effects on the striatum may help counter the dominant neural dynamic of the low arousal state, enabling a greater repertoire of potential neural activity patterns.

Alternatively, CL may drive the intralaminar influence over consciousness. Evidence from many studies suggests that 40-100Hz central thalamic stimulation induces conscious state transitions in rodents, macaques, or humans [100, 106, 107, 122]. Many neurons within CL present with a high, tonic, spontaneous firing rate (around 40-50Hz) during wakefulness,

which slows during general anesthesia and NREM sleep [107, 113]. It is possible that these CL neurons mediate the DBS rousing effect, with higher-frequency stimulation mimicking their wake-state firing patterns. Most of the DBS experiments covered in this review use clinical DBS electrodes, which have relatively large contacts and are likely to affect multiple regions by current spread. CM and CL are in close anatomical proximity, and thus most traditional DBS experiments stand a strong chance of influencing both nuclei with stimulation events. The use of a more-targeted, low current ($\leq$200μA) microstimulation approach has shown that emergence from anesthesia was more likely when the stimulation array was centered in CL, as opposed to neighboring nuclei like CM or MD [107]. Similar arousals observed in the other macaque experiments used substantially higher current (>1mA) targeted at CM, and weaker stimulation could not produce the same effects [108, 109]. This might suggest that higher current levels are needed to recruit activation of CL in order to restore consciousness.

One final but important clue about the underlying mechanism comes from a recent modelling study, where the CL-specific stimulation findings were recapitulated *in silico* using a biophysical neural mass model incorporating core and matrix thalamic projections. Simulated stimulations of the matrix-like, calbindin-rich thalamus (e.g., CL), but not core-like, parvalbumin-rich thalamus (e.g., CM), were shown to reinvigorate cortex and push the network towards a quasi-critical regime - where individual nodes in the network are highly susceptible to perturbations whilst also remaining stable at a network level - that is typically associated with wakefulness (Figure 2.2D). Through information theoretic analysis, this study showed that the shift back towards quasi-criticality allowed information to once again propagate across the network, restoring a coherent, wake-like state of bidirectional information flow despite the ongoing inhibitory pressure and reduced functional connectivity imposed by the simulated general anesthetic [123]. In spiking networks, analogous traveling waves have been shown to increase the response gain of individual neurons receiving input aligned with the waves phase and to gate perception of close-to-threshold visual stimuli [124, 125], thus providing a plausible link between the matrix-mediated changes in the state of consciousness discussed above and matrix-mediated gating of conscious content (a topic we discuss in detail in the next section). While there are limitations to such neural modelling experiments, the results suggest that regions like CL, with strong matrix-like projections to multiple cortical areas, are a better candidate to recruit cortical function than CM.

In sum, there is robust evidence that thalamus exerts vital control over conscious state. At the microscale, thalamus interfaces with the reticular activating system to tune $L5_{ET}$ neurons, controlling SWA and receptiveness of cortex to input. At the mesoscale, this translates into a broader mechanism by which the wakeful thalamus influences cortical functional connectivity and synchrony. The intralaminar nuclei, and perhaps CL in particular (with broad projections to both frontal and parietal cortex), may specifically control consciousness by tuning these cortical properties towards heightened conductivity at the macro scale. Overall then, while thalamus clearly plays a vital role in the arousal component of consciousness, the true power of its function seems to lie in the ability to reignite whole brain dynamics in ways that overcome inhibitory pressures of brain damage or anesthesia. In the rest of this review, we will make the case that thalamocortical circuits directly contribute to the contents of conscious experience.

2.4 Thalamocortical contributions to conscious contents

Although thalamocortical loops play a prominent role in several theories of consciousness [12, 74, 126], the search for the neural correlates of conscious content has, historically focused on the cerebral cortex (e.g., [127]). Rather than actively shaping experience, the thalamus (if it is considered at all) has typically been relegated to a background condition enabling conscious access [128]. Here, we advance the hypothesis, based on first-principles anatomical considerations, that matrix thalamic cells play a crucial computational role in shaping gating conscious contents. The diffuse connectivity of modulatory matrix cells [21, 69], combined with the widely branching apical dendrites of $L5_{ET}$'s [129] to which they project, seemingly does not possess adequate specificity to support the specificity of conscious contents. Rather, they offer a plausible mechanism for explaining how sensory content enters consciousness, irrespective of the properties of the specific content. Core thalamic neurons possess more precise driving connections to L2-4 IT-type neurons [21, 69], and interact with the inhibitory reticular nucleus which introduces a mechanism akin to divisive-normalization [130, 131] for selectivity into their dynamics. We argue that an appreciation of the nuanced neuroanatomy of the core (in addition to matrix) thalamic nuclei, and their embedding within the rest of the brain, may explain key properties of the contents of consciousness in typical waking experience. Although evidence is more sparse and the conclusions we can draw are therefore more speculative. We, therefore, focus on the role of matrix neurons in gating conscious contents. We will revisit the potential role of core thalamic neurons in determining the properties of conscious contents in the conclusion of the thesis.

2.4.1 Thalamus gates conscious contents via control of perceptual thresholds

The threshold for conscious perception is typically studied by contrasting neuronal responses to stimuli that are as close to physically-matched as possible but vary according to whether the subject is conscious of the stimulus (operationalised through direct or indirect behavioral reports [132, 133]). Recent work has keenly employed this approach in a mouse model of perception demonstrating a role for the same matrix thalamus-$L5_{ET}$ circuit implicated in the anesthetic control of conscious state [84] in controlling the threshold of conscious perception (Figures 2.2A and 2.3A [18, 134]).

In a whisker deflection threshold-detection task (Figures 2.3B-C), synchronous bursting (interspike interval < 12.5ms) in $L5_{ET}$'s was shown to be time-locked to stimulus onset (Figure 2.3D) and also to distinguish hits and false alarms (trials on which the animal reported perceiving a stimulus) from misses and correct rejections (trials on which the animal did not report perceiving the stimulus as present). As mentioned above, the bursting state of $L5_{ET}$'s *in vitro* is controlled by NMDAR-driven calcium spikes in apical dendrites, a prominent site of matrix thalamic inputs [22]. When there is a calcium spike in the apical dendrites and coincident input (within 25-30ms) to the basal dendrites, the somatic spikes of $L5_{ET}$ neurons shift from a regular spiking to a bursting regime [135]. Crucially, in the context of threshold detection [134], optogenetic stimulation and pharmacological inhibition of apical dendrites has been shown to shift the threshold for stimulus detection by making apical dendrites more or less likely to burst. Specifically, optogenetic activation of $L5_{ET}$ apical dendrites lowers the perceptual threshold of the animals, whilst a $GABA_B$ agonist increases the perceptual threshold [18]. Importantly, it was also shown that targeted inhibition of POm (a higher-order

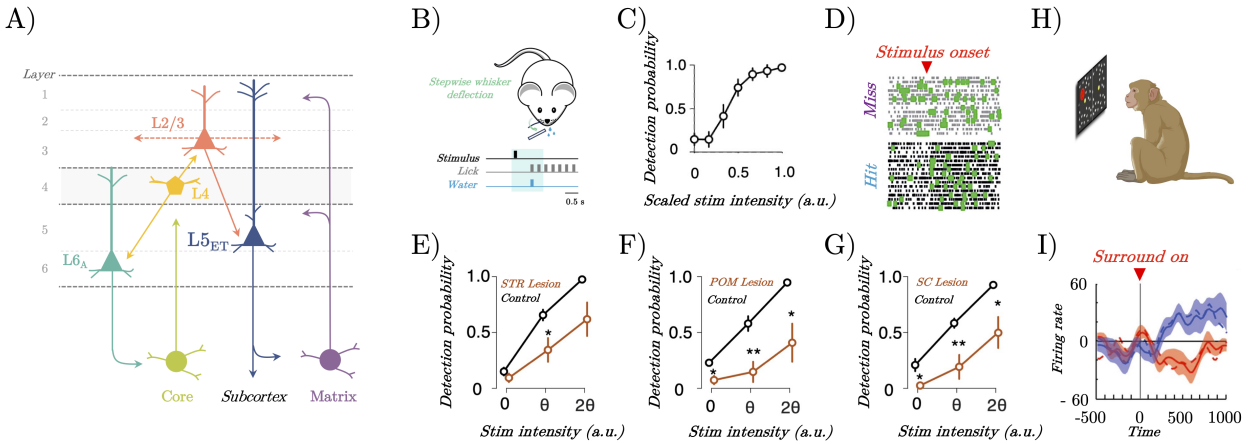

Figure 2.3: **A)** Thalamocortical circuit hypothesized to support the contents of consciousness. **B)** Illustration of the threshold detection paradigm. **C)** Psychometric function showing response probability. **D)** Example raster plots for the soma of $L5_{ET}$ cells on hit and miss trials. **E-G)** Detection probability as a function of stimulus intensity comparing control to inhibition of striatum, POm, or superior colliculus. **H)** Generalized continuous flash suppression paradigm. **I)** Average firing rate of neurons in the dorsal and ventral pulvinar.

somatosensory thalamic nucleus rich in both core and matrix cells), superior colliculus, and the striatum - all of which project either directly or indirectly to the apical dendrites of $L5_{ET}$ through matrix cells - increased the animals perceptual threshold (Figures 2.3E-G). Crucially, the synchronous burst response that was time-locked to stimulus onset was not found in a response/reward-only condition, suggesting that the synchronous bursting associated with conscious perception was unlikely to be driven by motor preparation or reward expectation. This implies that the threshold for the contents of conscious perception may be mediated by the formation of a reentrant loop between matrix thalamus and $L5_{ET}$ cells transitioning the cortical $L5_{ET}$ cells from a regular spiking (unconscious) to a bursting (conscious) regime [10, 15, 16].

Recent modelling work [136] has provided support for this perspective, showing that diffuse drive targeting the apical compartment of a large neuronal population of $L5_{ET}$'s - analogous to the projections of matrix thalamus - controlled the propagation of activity across the cortical sheet. Although the cellular-level study of the contents of consciousness in rodent models is still in its infancy, evidence from invasive recordings in human surgical patients [137] and non-human primates [138] as well as non-invasive human neuroimaging [139–141], combined with biophysical modelling [142, 143] and the anatomy of the matrix-$L5_{ET}$ circuit (Figure 2.3A), suggests that this circuit may be a key component in the general mechanisms underlying perceptual consciousness in humans, non-human primates, and rodents alike.

In humans, biophysical modelling of source-localized MEG recordings found that the auditory awareness negativity - an evoked electrophysiological response thought to characterize perceptual awareness of auditory stimuli in an auditory threshold detection task - was best explained by increased input to superficial layers of the cortical column housing the apical dendrites of $L5_{ET}$'s [143]. Given the known projections from matrix cells in higher-order thalamus to superficial layers of the cerebral cortex, a reasonable interpretation of this finding is that increased thalamic feedback in the aware condition led to widespread bursting in $L5_{ET}$'s. More direct evidence for thalamic involvement comes from a study of human patients with chronically-implanted thalamic electrodes [137]. Performance in a visual masking task revealed

a large, awareness-selective response in the matrix-rich intralaminar nuclei of the thalamus. Similar thalamic correlates of awareness, though coarser in grain, were confirmed using fMRI in healthy controls with a version of the task that controlled for report-related confounds. This suggests the thalamus is sensitive to perception above and beyond mechanisms needed to explicitly report the contents of perception. In non-human primates, single-unit recordings during generalized continuous flash suppression (Figure 2.3H) have shown that cells in ventral and dorsal pulvinar (higher-order thalamic nuclei, containing a mixture of matrix and core cells), are selectively modulated by an animals awareness of a stimulus, rather than its physical presence, independent of report demands (Figure 2.3I). In contrast, the first-order lateral geniculate nucleus (i.e., whose main driving input is from the retina) is selectively modulated by the physical presence of the stimulus, but not its subjective visibility [138]. This result is reinforced by three recent human neuroimaging studies of bistable perception, each of which found that visual rivalry induced activity in the pulvinar strongly differentiated between competing percepts [139–141]. This finding is also complemented by the result that lesions to pulvinar can lead to changes in conscious content, including perturbed feature binding [144], and hemineglect [145] which in the context of the current framework would manifest as a near absence of pulvinar-$L5_{ET}$ reentrant interaction preventing visual activity within the affected hemifield from achieving widespread propagation.

2.5 Chapter 2 discussion

We have argued that thalamocortical loops are a crucial component in the neuronal mechanisms controlling both conscious state and conscious contents, with distinct contributions from matrix and core thalamic cells. Specifically, there is growing evidence to suggest that matrix thalamic cells and their reciprocal, reentrant interactions with $L5_{ET}$ and $L6_{B}$ cortical neurons are vital for the desynchronized wake state [84]. Ultimately, the connections between these circuits (facilitated by orexin-sensitive $L6_{B}$ cells [17]) engage both broad (distributed) thalamo-cortical as well as local (intra-columnar) cortical circuitry, enabling the complex information processing modes required to support the wakeful conscious state.

In arguing for a thalamic contribution to the contents of consciousness, we drew on evidence showing that matrix-thalamus regulation of coupling between the apical and somatic compartments of $L5_{ET}$'s controls both conscious state as well as the threshold for the perception of conscious contents [18]. Due to the unique topological position of $L5_{ET}$'s in the cortical column, this coupling effectively decides whether activity can freely reverberate through cortico-cortical and thalamocortical loops, and thus achieve widespread influence throughout the brain, or whether it remains unconscious and isolated within an individual cortical column. A number of intralaminar nuclei, like CL, are not only predominantly constituted by matrix-type cells, they also have thalamostriatal projections which regulate basal ganglia inhibitory influence over thalamo-cortical loops. This provides an explanation of why the intralaminar thalamus holds a special place in shaping consciousness, and why it is well suited to DBS interventions in disorders of consciousness.

Based on the intimate connections between $L5_{ET}$'s and matrix thalamus in both the state and contents of consciousness, here we argued for a matrix-centric cellular explanation of conscious state and the gating of conscious contents that sits well with a number of previous theoretical proposals [15, 16, 70]. Substantiation of this conjecture requires more accurate identification of cell types and their link to perceptual experience. In addition, the pursuit of this goal of identifying correlates of conscious content down to the cellular level (whether

matrix, core, intermediate or other cell type), requires the use of experimental paradigms that can be run in parallel across human and model organisms alike [146]; animal models allow for high-resolution, high-density recording techniques from thalamo-cortical sites and causal interventions simply not possible in humans; whereas the lower resolution data in human participants can be used to directly validate the veracity of psychophysical methods used to study conscious experience. Perceptual rivalry variants represent an obvious candidate paradigm to use across species, as the phenomenon has been well characterized psychophysically in humans and non-human primates, does not rely on explicit report [147, 148], and has more recently been studied in mice [149].

In conclusion, taking into account the anatomy and functional evidence, it becomes impossible to relegate the thalamus to a mere background condition for consciousness. Close interactions between thalamus and cortex shape the neural landscape across micro, meso, and macro scales, giving rise to differences in information processing capacity and network interactions associated with different states. Mounting evidence suggests that these thalamocortical interactions directly shape the contents of consciousness by gating access to perception. Modern neuroscience tools make it possible to directly and precisely test these predictions. The time has come for detailed and dedicated study of the role of the thalamus and cortex, not as separate systems, but as a rich, integrated circuit that contributes integrally to both conscious state and conscious contents.

Chapter 2 references

[1] Wilfrid Sellars. "Philosophy and the scientific image of man". In: *Frontiers of science and philosophy* 1 (1962), pp. 1–40.

[2] Daniel C Dennett. *From bacteria to Bach and back: The evolution of minds*. WW Norton & Company, 2017.

[3] C. Klein and J. Hohwy. "Variability, convergence, and dimensions of consciousness". In: *Behavioural Methods in Consciousness*. 2015, pp. 249–263.

[4] S. Laureys. "The neural correlate of (un)awareness: lessons from the vegetative state". In: *Trends in Cognitive Sciences* 9 (2005), pp. 556–559.

[5] T. Bayne, J. Hohwy, and A.M. Owen. "Are There Levels of Consciousness?" In: *Trends in Cognitive Sciences* 20 (2016), pp. 405–413.

[6] A.K. Seth and T. Bayne. "Theories of Consciousness". In: *Nature Reviews Neuroscience* 23 (2022), pp. 439–452.

[7] I. Yaron et al. "The ConTraSt database for analysing and comparing empirical studies of consciousness theories". In: *Nature Human Behaviour* 6 (2022), pp. 593–604.

[8] D. Mumford. "On the Computational Architecture of the Neocortex. I. The Role of the Thalamo-Cortical Loop". In: *Biological Cybernetics* 65 (1991), pp. 135–145.

[9] L.M. Ward. "The thalamus: gateway to the mind". In: *Wiley Interdisciplinary Reviews: Cognitive Science* 4 (2013), pp. 609–622.

[10] J.M. Shine. "The Thalamus Integrates the Macrosystems of the Brain to Facilitate Complex, Adaptive Brain Network Dynamics". In: *Progress in Neurobiology* 199 (2021), p. 101951.

[11] M. Afrasiabi et al. "Consciousness depends on integration between parietal cortex, striatum, and thalamus". In: *Cell Systems* 12 (2021), 363–373.e11.

[12] G. Tononi and G.M. Edelman. "Consciousness and Complexity". In: *Science* 282 (1998), pp. 1846–1851.

[13] B.J. Baars. "Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience". In: *Progress in Brain Research*. Vol. 150. 2005, pp. 45–53.

[14] J.M. Shine et al. "The Impact of the Human Thalamus on Brain-Wide Information Processing". In: *Nature Reviews Neuroscience* 24 (2023), pp. 416–430.

[15] J. Aru, M. Suzuki, and M.E. Larkum. "Cellular Mechanisms of Conscious Processing". In: *Trends in Cognitive Sciences* 24 (2020), pp. 814–825.

[16] T. Bachmann, M. Suzuki, and J. Aru. "Dendritic integration theory: A thalamo-cortical theory of state and content of consciousness". In: *Philosophical Inquiries* 1 (2020), 10.33735/phimisci.2020.ii.52.

[17] T.A. Zolnik et al. "Layer 6b controls brain state via apical dendrites and the higher-order thalamocortical system". In: *Neuron* (2023). DOI: 10.1016/j.neuron.2023.11.021.

[18] N. Takahashi et al. "Active Dendritic Currents Gate Descending Cortical Outputs in Perception". In: *Nature Neuroscience* (2020). DOI: 10.1038/s41593-020-0677-8.

[19] E.G. Jones. "Viewpoint: the core and matrix of thalamic organization". In: *Neuroscience* 85 (1998), pp. 331–345.

[20] E.J. Müller et al. "Core and Matrix Thalamic Sub-Populations Relate to Spatio-Temporal Cortical Connectivity Gradients". In: *Neuroimage* 222 (2020), p. 117224.

[21] F. Clascá, P. Rubio-Garrido, and D. Jabaudon. "Unveiling the Diversity of Thalamocortical Neuron Subtypes". In: *European Journal of Neuroscience* 35 (2012), pp. 1524–1532.

[22] F. Brandalise et al. *Higher-order thalamocortical projections selectively control excitability via NMDAR and mGluRI-mediated mechanisms*. bioRxiv. 2023. DOI: 10.1101/2023.12.20.572353.

[23] S.M. Sherman. "Thalamus plays a central role in ongoing cortical functioning". In: *Nature Neuroscience* 19 (2016), pp. 533–541.

[24] S.M. Sherman and R.W. Guillery. "The Role of the Thalamus in the Flow of Information to the Cortex". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 357 (2002), pp. 1695–1708.

[25] J.W. Phillips et al. "A Repeated Molecular Architecture across Thalamic Pathways". In: *Nature Neuroscience* 22 (2019), pp. 1925–1935.

[26] E.G. Jones and S.H.C. Hendry. "Differential Calcium Binding Protein Immunoreactivity Distinguishes Classes of Relay Neurons in Monkey Thalamic Nuclei". In: *European Journal of Neuroscience* 1 (1989), pp. 222–246.

[27] M.C. Münkle, H.J. Waldvogel, and R.L. Faull. "The Distribution of Calbindin, Calretinin and Parvalbumin Immunoreactivity in the Human Thalamus". In: *Journal of Chemical Neuroanatomy* 19 (2000), pp. 155–173.

[28] E.G. Jones. "The Thalamic Matrix and Thalamocortical Synchrony". In: *Trends in Neurosciences* 24 (2001), pp. 595–601.

[29] E.G. Jones and R.Y. Leavitt. "Retrograde axonal transport and the demonstration of non-specific projections to the cerebral cortex and striatum from thalamic intralaminar nuclei in the rat, cat and monkey". In: *Journal of Comparative Neurology* 154 (1974), pp. 349–377.

[30] K.K. Cover and B.N. Mathur. "Rostral Intralaminar Thalamus Engagement in Cognition and Behavior". In: *Frontiers in Behavioral Neuroscience* 15 (2021), p. 652764.

[31] Y.D. Van der Werf, M.P. Witter, and H.J. Groenewegen. "The Intralaminar and Midline Nuclei of the Thalamus. Anatomical and Functional Evidence for Participation in Processes of Arousal and Awareness". In: *Brain Research Reviews* 39 (2002), pp. 107–140.

[32] A.M. Willis et al. "Open-Loop Organization of Thalamic Reticular Nucleus and Dorsal Thalamus: A Computational Model". In: *Journal of Neurophysiology* 114 (2015), pp. 2353–2367.

[33] D. Pinault. "The Thalamic Reticular Nucleus: Structure, Function and Concept". In: *Brain Research Reviews* 46 (2004), pp. 1–31.

[34] Y.J. John et al. "Visual Attention Deficits in Schizophrenia Can Arise From Inhibitory Dysfunction in Thalamus or Cortex". In: *Computational Psychiatry* 2 (2018), p. 223.

[35] J.M. Phillips, N.A. Kambi, and Y.B. Saalmann. "A Subcortical Pathway for Rapid, Goal-Driven, Attentional Filtering". In: *Trends in Neurosciences* 39 (2016), pp. 49–51.

[36] E.G. Jones. *The Thalamus*. Springer Science & Business Media, 2012.

[37] J.M. Shine. "Adaptively Navigating Affordance Landscapes: How Interactions between the Superior Colliculus and Thalamus Coordinate Complex, Adaptive Behaviour". In: *Neuroscience and Biobehavioral Reviews* 143 (2022), p. 104921.

[38] M. Nakajima and M.M. Halassa. "Thalamic Control of Functional Cortical Connectivity". In: *Current Opinion in Neurobiology* 44 (2017), pp. 127–131.

[39] S. Shipp. "The Functional Logic of Cortico-Pulvinar Connections". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358 (2003), pp. 1605–1624.

[40] J.M. Phillips et al. "Topographic organization of connections between prefrontal cortex and mediodorsal thalamus: Evidence for a general principle of indirect thalamic pathways between directly connected cortical areas". In: *Neuroimage* 189 (2019), pp. 832–846.

[41] J.M. Phillips et al. *Primate thalamic nuclei select abstract rules and shape prefrontal dynamics*. bioRxiv. 2024.

[42] J.M. Shine et al. "The Low-Dimensional Neural Architecture of Cognitive Complexity Is Related to Activity in Medial Thalamic Nuclei". In: *Neuron* 104 (2019), 849–855.e3.

[43] M.M. Halassa and Y.B. Saalmann. "The Mediodorsal Thalamus and Decision-Making". In: *The Cerebral Cortex and Thalamus*. Ed. by A.S.M.S.W. Martin Usrey. Oxford University Press, 2023.

[44] E.N. Brown, R. Lydic, and N.D. Schiff. "General anesthesia, sleep, and coma". In: *New England Journal of Medicine* 363 (2010), pp. 2638–2650.

[45] A. Destexhe, M. Rudolph, and D. Paré. "The High-Conductance State of Neocortical Neurons in Vivo". In: *Nature Reviews Neuroscience* 4 (2003), pp. 739–751.

[46] C.B. Saper et al. "Sleep state switching". In: *Neuron* 68 (2010), pp. 1023–1042.

[47] E.N. Brown, P.L. Purdon, and C.J. Van Dort. "General Anesthesia and Altered States of Arousal: A Systems Neuroscience Analysis". In: *Annual Review of Neuroscience* 34 (2011), pp. 601–628.

[48] P.M. Fuller et al. "Reassessment of the Structural Basis of the Ascending Arousal System". In: *Journal of Comparative Neurology* 519 (2011), pp. 933–956.

[49] C. Alexandre, M.L. Andermann, and T.E. Scammell. "Control of arousal by the orexin neurons". In: *Current Opinion in Neurobiology* 23 (2013), pp. 752–759.

[50] M.T. Alkire and J. Miller. "General anesthesia and the neural correlates of consciousness". In: *Progress in Brain Research*. Vol. 150. 2005, pp. 229–244.

[51] G.T. Neske. "The Slow Oscillation in Cortical and Thalamic Networks: Mechanisms and Functions". In: *Frontiers in Neural Circuits* 9 (2015), p. 88.

[52] M. Steriade. "Corticothalamic resonance, states of vigilance and mentation". In: *Neuroscience* 101 (2000), pp. 243–276.

[53] X.J. Wang. "Multiple dynamical modes of thalamic relay neurons: rhythmic bursting and intermittent phase-locking". In: *Neuroscience* 59 (1994), pp. 21–31.

[54] M. Deschênes et al. "Electrophysiology of neurons of lateral thalamic nuclei in cat: resting properties and burst discharges". In: *Journal of Neurophysiology* 51 (1984), pp. 1196–1219.

[55] A. Destexhe et al. "A model of spindle rhythmicity in the isolated thalamic reticular nucleus". In: *Journal of Neurophysiology* 72 (1994), pp. 803–818.

[56] F. David et al. "Essential thalamic contribution to slow waves of natural sleep". In: *The Journal of Neuroscience* 33 (2013), pp. 19599–19610.

[57] M. Lemieux et al. "The impact of cortical deafferentation on the neocortical slow oscillation". In: *The Journal of Neuroscience* 34 (2014), pp. 5689–5703.

[58] Y.B. Saalmann. "Intralaminar and medial thalamic influence on cortical synchrony, information transmission and cognition". In: *Frontiers in Systems Neuroscience* 8 (2014), p. 83.

[59] Y.B. Saalmann and S. Kastner. "Cognitive and perceptual functions of the visual thalamus". In: *Neuron* 71 (2011), pp. 209–223.

[60] B. Setzer et al. “A temporal sequence of thalamic activity unfolds at transitions in behavioral arousal state”. In: *Nature Communications* 13 (2022), p. 5442.
[61] W. Singer. “Consciousness and neuronal synchronization”. In: *The neurology of consciousness*. 2011, pp. 43–52.
[62] A.K. Engel and P. Fries. “Neuronal Oscillations, Coherence, and Consciousness”. In: *The Neurology of Conciousness (Second Edition)*. Ed. by S. Laureys, O. Gosseries, and G. Tononi. Academic Press, 2016, pp. 49–60.
[63] S. Casarotto et al. “Stratification of unresponsive patients by an independently validated index of brain complexity”. In: *Annals of Neurology* 80 (2016), pp. 718–729.
[64] F. Ferrarelli et al. “Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness”. In: *Proceedings of the National Academy of Sciences of the United States of America* 107 (2010), pp. 2681–2686.
[65] M. Napolitani et al. “Transcranial magnetic stimulation combined with high-density EEG in altered states of consciousness”. In: *Brain Injury* 28 (2014), pp. 1180–1189.
[66] S. Sarasso et al. “Consciousness and Complexity during Unresponsiveness Induced by Propofol, Xenon, and Ketamine”. In: *Current Biology* 25 (2015), pp. 3099–3105.
[67] L.D. Claar et al. “Cortico-thalamo-cortical interactions modulate electrically evoked EEG responses in mice”. In: *eLife* 12 (2023). DOI: 10.7554/eLife.84630.
[68] M.L. Cavelli et al. “Sleep/wake changes in perturbational complexity in rats and mice”. In: *iScience* 26 (2023), p. 106186.
[69] K.D. Harris and G.M.G. Shepherd. “The neocortical circuit: themes and variations”. In: *Nature Neuroscience* 18 (2015), pp. 170–181.
[70] J. Aru et al. “Coupling the State and Contents of Consciousness”. In: *Frontiers in Systems Neuroscience* 13 (2019), p. 43.
[71] R.J. Douglas and K.A.C. Martin. “Neuronal circuits of the neocortex”. In: *Annual Review of Neuroscience* 27 (2004), pp. 419–451.
[72] K. Friston and S. Kiebel. “Predictive coding under the free-energy principle”. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 364 (2009), pp. 1211–1221.
[73] N.T. Markov et al. “Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex”. In: *Journal of Comparative Neurology* 522 (2014), pp. 225–259.
[74] G.A. Mashour et al. “Conscious Processing and the Global Neuronal Workspace Hypothesis”. In: *Neuron* 105 (2020), pp. 776–798.
[75] J.F. Mejias et al. “Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex”. In: *Science Advances* 2 (2016), e1601335.
[76] S.R. Olsen et al. “Gain control by layer six in cortical circuits of vision”. In: *Nature* 483 (2012), pp. 47–52.
[77] A.M. Thomson and C. Lamy. “Functional maps of neocortical local circuitry”. In: *Frontiers in Neuroscience* 1 (2007), pp. 19–42.
[78] J. Voigts, C.A. Deister, and C.I. Moore. “Layer 6 ensembles can selectively regulate the behavioral impact and layer-specific representation of sensory deviants”. In: *eLife* 9 (2020). DOI: 10.7554/eLife.48957.
[79] R. Beltramo et al. “Layer-specific excitatory circuits differentially control recurrent network dynamics in the neocortex”. In: *Nature Neuroscience* 16 (2013), pp. 227–234.
[80] A. Stroh et al. “Making waves: initiation and propagation of corticothalamic Ca2+ waves in vivo”. In: *Neuron* 77 (2013), pp. 1136–1150.

[81] M.V. Sanchez-Vives. "Origin and dynamics of cortical slow oscillations". In: *Current Opinion in Physiology* 15 (2020), pp. 217–223.
[82] A. Bharioke et al. "General anesthesia globally synchronizes activity selectively in layer 5 cortical pyramidal neurons". In: *Neuron* 110 (2022), 2024–2040.e10.
[83] G. Buzsaki. *Rhythms of the Brain*. Oxford University Press, 2006.
[84] M. Suzuki and M.E. Larkum. "General Anesthesia Decouples Cortical Pyramidal Neurons". In: *Cell* 180 (2020), 666–676.e13.
[85] R.W. Guillery. "Anatomical evidence concerning the role of the thalamus in corticocortical communication: a brief review". In: *Journal of Anatomy* 187 (1995), pp. 583–592.
[86] J.M. Phillips et al. "Disentangling the Influences of Multiple Thalamic Nuclei on Prefrontal Cortex and Cognitive Control". In: *Neuroscience and Biobehavioral Reviews* 128 (2021), pp. 487–510.
[87] Y.B. Saalmann et al. "The pulvinar regulates information transmission between cortical areas based on attention demands". In: *Science* 337 (2012), pp. 753–756.
[88] Y. Nir et al. "Regional slow waves and spindles in human sleep". In: *Neuron* 70 (2011), pp. 153–169.
[89] D. Miyamoto, D. Hirai, and M. Murayama. "The Roles of Cortical Slow Waves in Synaptic Plasticity and Memory Consolidation". In: *Frontiers in Neural Circuits* 11 (2017), p. 92.
[90] M. Boly et al. "Connectivity changes underlying spectral EEG changes during propofol-induced loss of consciousness". In: *The Journal of Neuroscience* 32 (2012), pp. 7082–7090.
[91] L.D. Lewis et al. "Rapid fragmentation of neuronal networks at the onset of propofol-induced unconsciousness". In: *Proceedings of the National Academy of Sciences of the United States of America* 109 (2012), E3377–E3386.
[92] N.D. Schiff. "Central thalamic contributions to arousal regulation and neurological disorders of consciousness". In: *Annals of the New York Academy of Sciences* 1129 (2008), pp. 105–118.
[93] B.L. Edlow et al. "Disconnection of the ascending arousal system in traumatic coma". In: *Journal of Neuropathology and Experimental Neurology* 72 (2013), pp. 505–523.
[94] J. Hindman et al. "Thalamic strokes that severely impair arousal extend into the brainstem". In: *Annals of Neurology* 84 (2018), pp. 926–930.
[95] B.L. Edlow et al. "Recovery from disorders of consciousness: mechanisms, prognosis and emerging therapies". In: *Nature Reviews Neurology* 17 (2021), pp. 135–156.
[96] U.E. Heinz and J.D. Rollnik. "Outcome and prognosis of hypoxic brain damage patients undergoing neurological early rehabilitation". In: *BMC Research Notes* 8 (2015), p. 243.
[97] S. Laureys, M. Boly, and P. Maquet. "Tracking the recovery of consciousness from coma". In: *Journal of Clinical Investigation* 116 (2006), pp. 1823–1825.
[98] S. Laureys et al. "Restoration of thalamocortical connectivity after recovery from persistent vegetative state". In: *The Lancet* 355 (2000), pp. 1790–1791.
[99] P. Pais-Roldán et al. "Multimodal assessment of recovery from coma in a rat model of diffuse brainstem tegmentum injury". In: *Neuroimage* 189 (2019), pp. 615–630.
[100] N.D. Schiff et al. "Behavioural improvements with thalamic stimulation after severe traumatic brain injury". In: *Nature* 448 (2007), pp. 600–603.
[101] J.A. Cain et al. "Real time and delayed effects of subcortical low intensity focused ultrasound". In: *Scientific Reports* 11 (2021), p. 6100.

[102] D. Chudy et al. “Deep brain stimulation in disorders of consciousness: 10 years of a single center experience”. In: *Scientific Reports* 13 (2023), p. 19491.
[103] H. Arnts et al. “Clinical and neurophysiological effects of central thalamic deep brain stimulation in the minimally conscious state after severe brain injury”. In: *Scientific Reports* 12 (2022), p. 12932.
[104] T. Cao et al. “Clinical neuromodulatory effects of deep brain stimulation in disorder of consciousness: A literature review”. In: *CNS Neuroscience and Therapeutics* (2023). DOI: 10.1111/cns.14559.
[105] F. Velasco et al. “Role of the centromedian thalamic nucleus in the genesis, propagation and arrest of epileptic activity. An electrophysiological study in man”. In: *Acta Neurochirurgica Supplement*. Vol. 58. 1993, pp. 201–204.
[106] M. Velasco et al. “Electrocortical and behavioral responses produced by acute electrical stimulation of the human centromedian thalamic nucleus”. In: *Electroencephalography and Clinical Neurophysiology* 102 (1997), pp. 461–471.
[107] M.J. Redinbaugh et al. “Thalamus Modulates Consciousness via Layer-Specific Control of Cortex”. In: *Neuron* 106 (2020), 66–75.e12.
[108] A.M. Bastos et al. “Neural effects of propofol-induced unconsciousness and its reversal using thalamic stimulation”. In: *eLife* 10 (2021). DOI: 10.7554/eLife.60824.
[109] J. Tasserie et al. “Deep brain stimulation of the thalamus restores signatures of consciousness in a nonhuman primate model”. In: *Science Advances* 8 (2022), eabl5547.
[110] E.J. Tehovnik. “Electrical stimulation of neural tissue to evoke behavioral responses”. In: *Journal of Neuroscience Methods* 65 (1996), pp. 1–17.
[111] S. Patra. “Response properties of human thalamic neurons to high frequency microstimulation”. In: (2001).
[112] M.J. Redinbaugh et al. “Thalamic deep brain stimulation paradigm to reduce consciousness: Cortico-striatal dynamics implicated in mechanisms of consciousness”. In: *PLoS Computational Biology* 18 (2022), e1010294.
[113] L.L. Glenn and M. Steriade. “Discharge rate and excitability of cortically projecting intralaminar thalamic neurons during waking and sleep states”. In: *The Journal of Neuroscience* 2 (1982), pp. 1387–1404.
[114] L.C. Towns, J. Tigges, and M. Tigges. “Termination of thalamic intralaminar nuclei afferents in visual cortex of squirrel monkey”. In: *Visual Neuroscience* 5 (1990), pp. 151–154.
[115] E.F. Kaufman and A.C. Rosenquist. “Efferent projections of the thalamic intralaminar nuclei in the cat”. In: *Brain Research* 335 (1985), pp. 257–279.
[116] M. Parent, M. Lévesque, and A. Parent. “Two types of projection neurons in the internal pallidum of primates: single-axon tracing and three-dimensional reconstruction”. In: *Journal of Comparative Neurology* 439 (2001), pp. 162–175.
[117] D.L. Kasdon and S. Jacobson. “The thalamic afferents to the inferior parietal lobule of the rhesus monkey”. In: *Journal of Comparative Neurology* 177 (1978), pp. 685–706.
[118] J. Kievit and H.G. Kuypers. “Subcortical afferents to the frontal lobe in the rhesus monkey studied by means of retrograde horseradish peroxidase transport”. In: *Brain Research* 85 (1975), pp. 261–266.
[119] B.A. Vogt, D.L. Rosene, and D.N. Pandya. “Thalamic and cortical afferents differentiate anterior from posterior cingulate cortex in the monkey”. In: *Science* 204 (1979), pp. 205–207.
[120] N.D. Schiff. “Recovery of consciousness after brain injury: a mesocircuit hypothesis”. In: *Trends in Neurosciences* 33 (2010), pp. 1–9.

[121] T. Minamimoto, Y. Hori, and M. Kimura. "Complementary process to response bias in the centromedian nucleus of the thalamus". In: *Science* 308 (2005), pp. 1798–1801.
[122] J. Liu et al. "Frequency-selective control of cortical and subcortical networks by central thalamus". In: *eLife* 4 (2015), e09215.
[123] E.J. Müller et al. "The Non-Specific Matrix Thalamus Facilitates the Cortical Information Processing Modes Relevant for Conscious Awareness". In: *Cell Reports* 42 (2023), p. 112844.
[124] Z.W. Davis et al. "Spontaneous travelling cortical waves gate perception in behaving primates". In: *Nature* 587 (2020), pp. 432–436.
[125] Z.W. Davis et al. "Spontaneous traveling waves naturally emerge from horizontal fiber time delays and travel through locally asynchronous-irregular states". In: *Nature Communications* 12 (2021), p. 6057.
[126] M. Oizumi, L. Albantakis, and G. Tononi. "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0". In: *PLoS Computational Biology* 10 (2014), e1003588.
[127] S. Bisenius et al. "Identifying neural correlates of visual consciousness with ALE meta-analyses". In: *Neuroimage* 122 (2015), pp. 177–187.
[128] S. Dehaene and J.-P. Changeux. "Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness". In: *PLoS Biology* 3 (2005), e141.
[129] S. Ramaswamy and H. Markram. "Anatomy and physiology of the thick-tufted layer 5 pyramidal neuron". In: *Frontiers in Cellular Neuroscience* 9 (2015), p. 233.
[130] Y.J. John et al. "The Emotional Gatekeeper: A Computational Model of Attentional Selection and Suppression through the Pathway from the Amygdala to the Inhibitory Thalamic Reticular Nucleus". In: *PLoS Computational Biology* 12 (2016), e1004722.
[131] S. Grossberg. "A neural model of attention, reinforcement and discrimination learning". In: *International Review of Neurobiology* 18 (1975), pp. 263–327.
[132] J. Aru et al. "Distilling the neural correlates of consciousness". In: *Neuroscience and Biobehavioral Reviews* 36 (2012), pp. 737–746.
[133] M. Overgaard. *Behavioural Methods in Consciousness Research*. Oxford University Press, 2015.
[134] N. Takahashi et al. "Active Cortical Dendrites Modulate Perception". In: *Science* 354 (2016), pp. 1587–1590.
[135] M.E. Larkum. "Top-down Dendritic Input Increases the Gain of Layer 5 Pyramidal Neurons". In: *Cerebral Cortex* 14 (2004), pp. 1059–1070.
[136] B.R. Munn et al. "Neuronal connected burst cascades bridge macroscale adaptive signatures across arousal states". In: *Nature Communications* 14 (2023), p. 6846.
[137] S.I. Kronemer et al. "Human visual consciousness involves large scale cortical and subcortical networks independent of task report and eye movement activity". In: *Nature Communications* 13 (2022), p. 7342.
[138] M. Wilke, K.-M. Mueller, and D.A. Leopold. "Neural activity in the visual thalamus reflects perceptual suppression". In: *Proceedings of the National Academy of Sciences of the United States of America* 106 (2009), pp. 9465–9470.
[139] C. Qian et al. *Hierarchical and fine-scale mechanisms of binocular rivalry for conscious perception*. bioRxiv. 2023. DOI: 10.1101/2023.02.11.528110.
[140] J.W. Kurzawski et al. "Short-term plasticity in the human visual thalamus". In: *eLife* 11 (2022). DOI: 10.7554/eLife.74565.

[141] J. Seo et al. “The thalamocortical inhibitory network controls human conscious perception”. In: *Neuroimage* 264 (2022), p. 119748.

[142] C.J. Whyte et al. *A Burst-dependent Thalamocortical Substrate for Visual Rivalry*. bioRxiv. 2023. DOI: 10.1101/2023.07.13.548934.

[143] C. Fernandez Pujol, E.G. Blundon, and A.R. Dykstra. “Laminar specificity of the auditory perceptual awareness negativity: A biophysical modeling study”. In: *PLoS Computational Biology* 19 (2023), e1011003.

[144] R. Ward et al. “Deficits in spatial coding and feature binding following damage to spatiotopic maps in the human pulvinar”. In: *Nature Neuroscience* 5 (2002), pp. 99–100.

[145] H.O. Karnath, M. Himmelbach, and C. Rorden. “The subcortical anatomy of human spatial neglect: putamen, caudate nucleus and pulvinar”. In: *Brain* 125 (2002), pp. 350–360.

[146] B.J. He. “Towards a pluralistic neurobiological understanding of consciousness”. In: *Trends in Cognitive Sciences* 27 (2023), pp. 420–432.

[147] S. Frässle et al. “Binocular rivalry: frontal activity relates to introspection and action but not to perception”. In: *The Journal of Neuroscience* 34 (2014), pp. 1738–1747.

[148] V. Kapoor et al. “Decoding internally generated transitions of conscious contents in the prefrontal cortex without subjective reports”. In: *Nature Communications* 13 (2022), p. 1535.

[149] G. Palagina, J.F. Meyer, and S.M. Smirnakis. “Complex Visual Motion Representation in Mouse Area V1”. In: *The Journal of Neuroscience* 37 (2017), pp. 164–183.

Chapter 3

A Burst-dependent Thalamocortical Substrate for Perceptual Awareness

> Nature is exceedingly simple and comfortable to herself. Whatever reasoning holds for greater motions should hold for lesser ones as well.
>
> — *Isaac Newton*

In the previous chapter, we argued that the same matrix thalamus - $L5_{ET}$ circuit motif that underlies the global state of consciousness also controls the threshold for perception, gating the perceptual contents of awareness. The primary evidence for this claim came from a mouse model of threshold detection [1, 2] and was indirectly bolstered by data from human neuroimaging [3, 4] and non-human primate electrophysiology [5] showing visibility-specific activity in high-order visual thalamus. In this chapter, I develop and extend this hypothesis *in silico* through microscale spiking neural network simulations. Although my ultimate aim is to develop a cross-level mechanistic theory [6] of perceptual awareness, with models implemented at multiple scales and levels of analysis, in order to implement any singular model, you have to pick a scale. To facilitate comparison between simulations and empirical data, I chose to start at the spiking level, as this is the scale closest to the neural recordings. I am working under the Newtonian assumption that principles and lessons gleaned at this 'lesser' scale will generalise to 'greater' scales.

This chapter is the centrepiece of the thesis; it links microscale cellular-level mechanisms, discovered and characterised in a mouse model of perception, to behavioural signatures of awareness studied in human psychophysics. Developing the model took the better part of two years, and I rebuilt it from the ground up on at least five separate occasions. This process, although painful at times, was well worth it. The final product is exceptionally robust, with results that generalise over a wide range of parameter values and network sizes.

3.1 Chapter 3 introduction

As I argued in the introduction, the study of perceptual awareness — the process of gaining conscious access to perceptual content — in human participants (e.g. [7–9]) and animal models (e.g. [10–13]) has opposing but complementary limitations. Both fields contain crucial pieces of the puzzle for understanding perceptual awareness; however the links between the two are currently limited at best. Effective progress, therefore, hinges on our ability to create empirically tractable tethers between the behavioural signatures of perceptual awareness studied in humans and the fine-grained neurobiological mechanisms studied in animal models [14].

Recent work in a mouse model of perception has identified a key thalamocortical circuit connecting thick-tufted layer 5 extratelencephalic ($L5_{ET}$) neurons and matrix thalamic cells as playing a causal role in the threshold for perceptual awareness [15–19]. Specifically, based on a range of cellular recordings and causal perturbations, it has been shown that matrix-thalamus-mediated coupling of apical dendrite and somatic compartments in $L5_{ET}$ cells leads to a burst-firing state that is a reliable signature of perceptual awareness of a near-threshold tactile stimulus [18, 19]. However, the simplicity of the threshold detection task and species-specific differences in neural architecture means that it is not clear whether the mechanisms of perceptual awareness characterised in the mouse model will generalise beyond the whisker detection task to the more complex paradigms typically studied in human participants. Here, we use biophysical modelling to bridge the gap between the thalamocortical circuit identified in the mouse model of perception [16, 18, 19] and the behavioural signatures of perceptual awareness studied in human psychophysics. Specifically, we built a thalamocortical spiking neural network model that explains the full suite of behavioural and neural findings in the mouse model of tactile threshold detection.

Given the ubiquity of the thalamocortical circuit architecture across sensory modalities, we [16, 17, 20], along with others [21, 22], have proposed that reverberant bursting activity in $L5_{ET}$ - matrix thalamus loops may be a necessary component part in a domain-general mechanism of perceptual awareness. A key test of this hypothesis is whether this same circuit architecture can explain psychophysical principles known to govern perceptual awareness in more complex paradigms and in other sensory modalities. To test this hypothesis *in silico*, we leveraged the same model with identical parameters to simulate both tactile threshold detection and visual rivalry (which we use as a catch-all term for binocular rivalry and related bistable perception paradigms).

Visual rivalry is a complex but highly psychophysically-constrained phenomenon whereby visual perception stochastically switches between stimulus percepts that differ only in terms of their perceptual content [23–26]. Visual rivalry provides a means to dissociate the neural mechanisms of subjective perception from the correlates of physical stimulation. Crucially, variants of visual rivalry (i.e., plaid rivalry; [27]) can now be studied in mouse models [13, 28] as well in human and non-human primates, allowing us to test our model against existing psychophysical findings in primates and to make predictions for what should be observed in the mouse model before data has been collected - a central component in the evolving dialogue between theory and experiment. In addition, although our model is consistent with the psychophysical predictions of previous models of visual rivalry (e.g., [29–37]), our approach has an unique level of neurobiological specificity that allows us to generate cellular level predictions about the neural underpinnings of perceptual awareness in a language that is

applicable to the causal methods used by modern systems neuroscientists.

3.2 Chapter 3 results

3.2.1 A spiking corticothalamic model recreates key features of cellular physiology

The dynamical elements of the model were inspired by recent empirical observations, and consist of three classes of neurons (Figure 3.1A-B) - $L5_{ET}$ cells (blue), fast spiking interneurons (basket cells, gold), and diffuse projecting matrix thalamocortical cells (purple) - each of which are modelled using biophysically plausible spiking neurons [38–41] coupled through conductance-based synapses. By building the model from these circuit elements, we ensured that the emergent dynamics of the population recapitulated known signatures of cell-type-specific firing patterns, thus retaining the capacity to translate insights between computational modellers and cellular physiologists.

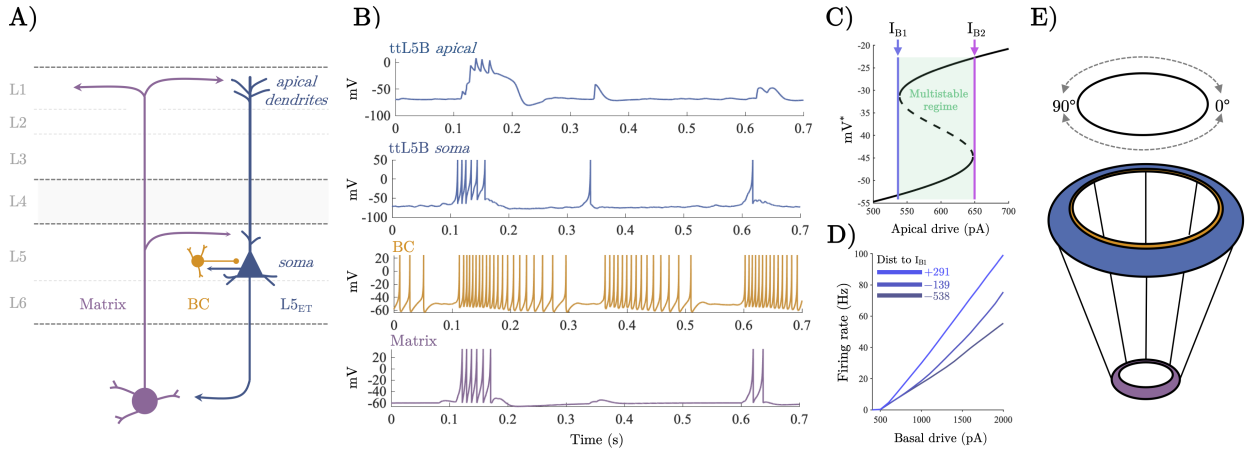

Figure 3.1: **A)** Idealised anatomy of the model thalamocortical loop connecting higher-order matrix thalamus and $L5_{ET}$ neurons. **B)** Single neuron dynamics of example neurons in the thalamocortical network for each class of cell when driven with 600 Hz of independent background drive to the somatic compartment of every neuron and 50 Hz to apical compartment of $L5_{ET}$ cells. **C)** Bifurcation diagram of the $L5_{ET}$ apical compartment. I_{B1} denotes a saddle node bifurcation generating a stable plateau potential which coexists with the resting state of the apical compartment until the model passes through a second saddle node bifurcation at I_{B2} at which point the resting state vanishes and the stable plateau potential becomes globally attracting. **D)** Somatic firing rate of the novel dual compartment $L5_{ET}$ model as a function of basal (i.e. somatic compartment) drive and apical drive (measured in terms of the distance to I_{B1}). In line with empirical data [42], apical drive increases the gain of the somatic compartment. **E)** Model thalamocortical ring architecture. $L5_{ET}$ cells and basket cells were placed on a cortical ring at evenly space intervals.

To model the non-linear bursting behaviour of $L5_{ET}$ cells (which has been linked to perceptual awareness; [42–44]), we created a novel dual-compartment model with active apical dendrites that captures the essential features of the cells' physiology. Empirical recordings have shown that $L5_{ET}$ cells switch from regular spiking to bursting when they receive near-simultaneous input to both their apical (top) and basal (bottom) dendrites (Figure 3.1B; [42, 43]). As such, our model includes a somatic compartment, described by an Izhikevich adaptive quadratic integrate and fire neuron [40, 45, 46] and an apical compartment, described by a non-linear model of the Ca^{2+} plateau potential [41, 47]. To recapitulate known physiology, we coupled the compartments such that sodium spikes in the somatic compartment back propagate to the

apical compartment; in turn if a Ca^{2+} plateau potential is triggered in the apical compartment the somatic compartment's behaviour (probabilistically) switches from regular spiking to bursting (implemented by switching the reset conditions of the somatic compartment). The amount of current entering the apical compartment controls this switching process. With sufficiently high current, the apical compartment enters inter a multistable regime marked by the saddle node bifurcation at I_{B1}(Figure 3.1C), where a stable Ca^{2+} plateau potential can coexist with the resting state of the compartment. Further increases in current cause the cell's resting state to disappear by passing the cell through a second saddle node bifurcation at I_{B2} (Figure 3.1C), making the plateau potential globally attracting (see supplementary material 3.5.1). Based on the finding that communication between the soma and apical dendrites of $L5_{ET}$ cells depends upon depolarising input from the matrix thalamus to the apical coupling zone of $L5_{ET}$ cells [48], we made the probability of successful propagation between compartments proportional to the amplitude of thalamic conductances. In line with empirical findings and previous modelling [42, 49], Ca^{2+} plateau potentials in the apical compartment controlled the gain of the somatic compartment's firing rate curve by increasing the amount of time the somatic compartment spent in a bursting rather than a regular spiking parameter regime (Figure 3.1D).

In line with previous spiking neural network models of early sensory cortex [30, 31, 50, 51], we embedded the cortical neurons in a one-dimensional ring architecture (90 pairs of $L5_{ET}$ excitatory and fast-spiking inhibitory interneurons; Figure 3.1E). Each point on the ring represents an orientation preference, with one full rotation around the ring corresponding to a 180° visual rotation - this provides each neuron with a 2° difference in orientation preference, relative to its neighbours. The cortical ring was coupled to a thalamic ring with a 9:1 ratio (to approximately reflect the cortico-thalamic ratio in mammals), which then projected back up to the apical dendrites of the same 9 cortical neurons, representing the diffuse projections of higher-order thalamus onto the apical dendrites of $L5_{ET}$ neurons in layer 1 [52, 53]. Cortical coupling was modelled with a spatial decay, with long range inhibitory coupling and comparatively local excitatory coupling (i.e., centre-surround Mexican-hat' connectivity). When driven solely by baseline input, the model emitted irregular spikes interspersed with sparse spatially localised bursts mediated by depolarising input from the thalamus which allowed $L5_{ET}$ somatic spikes to back-propagate initiating Ca^{2+} plateau potentials in the apical dendrites which in turn initiated transient burst spiking in the somatic compartment (Figure 3.1C). Cortical spiking activity was highly irregular (mean inter-spike interval coefficient of variation standard deviation/mean = 2.4), characteristic of a waking state [54]. For more details on the model architecture and analysis of the dynamics, see the methods section.

3.2.2 Thalamocortical model reproduces empirical signatures of threshold detection

We first set out to reproduce the results of the whisker-based tactile detection paradigm employed by Takahashi and colleagues [18, 19], who trained mice to report a mechanical deflection of a whisker over a range of deflection intensities (Figure 3.2A) while recording $L5_{ET}$ activity in barrel cortex from the apical dendrites via fast scanning two-photon Ca^{2+} imaging, and somatic activity via juxtacellular electrodes [18, 19]. The original study found that bursting activity in the soma of $L5_{ET}$ cells, generated by Ca^{2+} plateau potentials in the apical dendrites, distinguished hits and false alarms from misses and correct rejections. Importantly, they were able to establish causality through a series of perturbation experiments (Figure 3.2B). Optogenetic excitation of the apical dendrites reduced the animal's threshold for awareness increasing both hits and false-alarms (Figure 3.2C). In turn, pharmocological inhibition of the apical dendites and POm (a matrix-rich higher-order thalamic nucleus with closed loop connections to barrel cortex; [52]) increased the animals perceptual threshold (Figure 3.2D-E).

To model the perceptual discrimination process underlying threshold detection - discriminating the presence of a weak stimulus against a noisy background - all neurons received a constant background drive consisting of independent Poisson spike trains while an arbitrary cortical neuron was pulsed by a current of constant width and variable amplitude that we weighted by a spatial Gaussian to mimic the selectivity of neurons in early sensory cortex. We operationalised perceptual awareness in the threshold detection simulations in terms of what an upstream ideal observer could readout from the population by computing whether trial-by-trial spike counts in the 1000 ms post stimulus window exceeded an optimal criterion (i.e. the criterion that best minimises misses and false alarms across stimulus intensities). We counted the model as having made a response whenever the spike count exceeded the optimal criterion. Psychometric functions were then fit to the model's responses (for details see the methods section). Qualitatively identical results were obtained using neurometric functions which sum over all criterion values (see supplementary material 3.5.2; [55]).

The model's responses to simulated whisker deflections recapitulated the empirically observed sigmoid-like relationship between the intensity of the simulation and the model's response probability (Figure 3.2F-H). In addition, perturbations to the model designed to replicate optogenetic excitation and pharmacological inhibition (Figure 3.2A) qualitatively reproduced the empirically observed shifts in response probability across perturbation types (Figure 3.2C-H). For each type of perturbation we ran the simulation over a range of perturbation magnitudes to ensure the reliability of the effect (supplementary material 3.5.2). For brevity, we only show the results for 300 pA perturbations in the main text. In addition, we note that the saturating detection probability values in the model are due to an absense of behavioural stochasticity resulting from extranious factors such as decision noise which are inherent to empirical data.

A key benefit of biophysical modelling is the capacity to mechanistically probe the model and determine how the empirical observations may have emerged from the underlying circuit dynamics. To this end, we used tools from dynamical systems theory to interrogate the cell-type-specific dynamics underlying the behaviour of the model including, the distance to bifurcation (I_{B1}) in the apical compartment, the parameter regime of the somatic compartment (i.e. regular spiking or bursting), and the inter-compartment coupling probability. Excitation of the apical dendrites (blue) reduced the average distance to bifurcation (here defined as the distance to I_{B1}) in the apical compartment across the network in the 1000 ms period post stim-

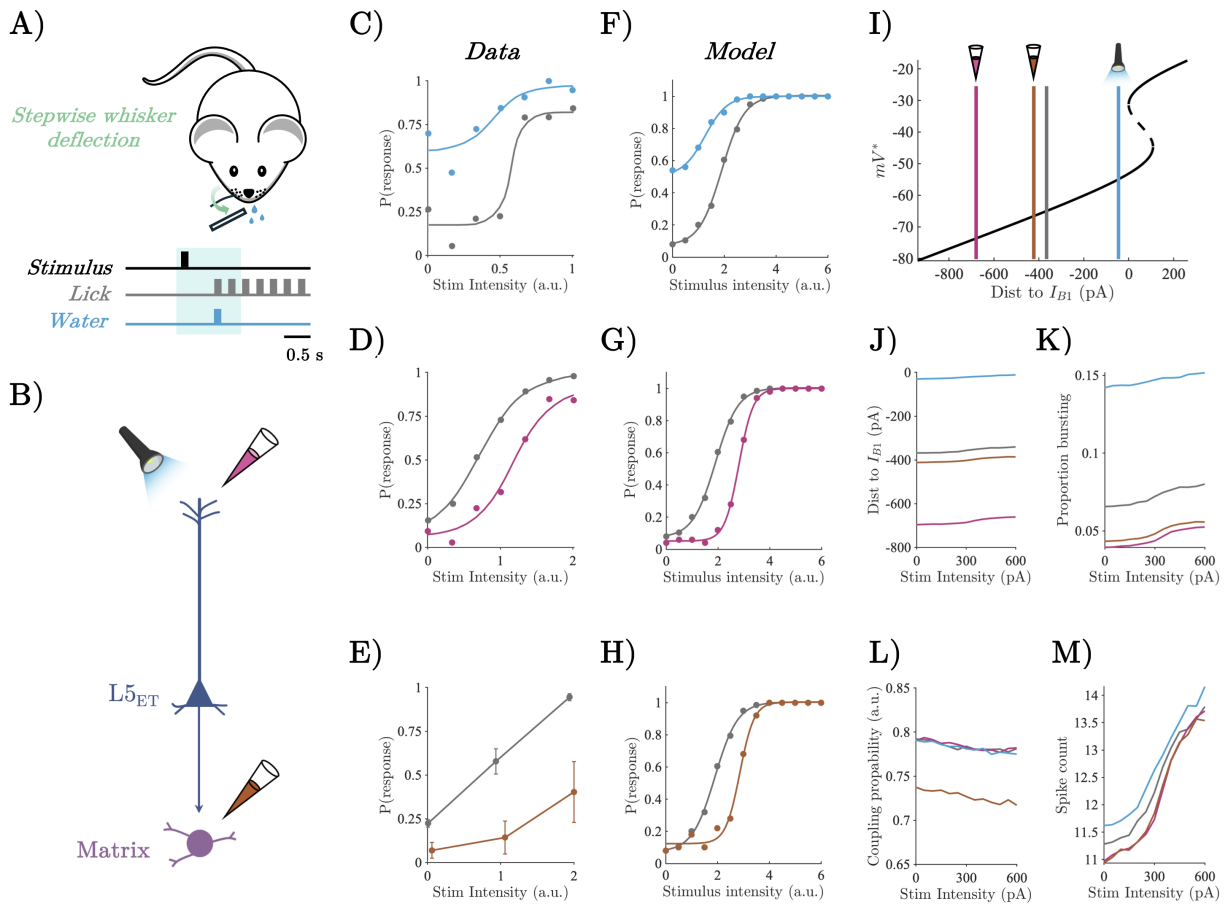

Figure 3.2: **A)** Whisker deflection paradigm of [19]. **B)** Representation of causal perturbations to $L5_{ET}$ thalamocortical circuit including optogenetic excitation of the apical dendrites (light blue), pharmacological inhibition of the apical dendrites (pink), and pharmacological inhibition of POm (orange). **C-D)** Psychometric function of animals performing a whisker deflection task for the control condition (grey), optogenetic excitation of apical dendrites (blue), and pharmacological inhibition of apical dendrites (pink). Modified from [18] and [19]. **E)** Response probability as a function of whisker deflection intensity for control (grey) and pharmacological inhibition of POm (orange). Modified from [19]. **F-H)** Model response probability as a function of stimulus intensity across simulated causal perturbations. Colours same as above. **I)** Apical compartment bifurcation diagram showing the average distance to I_{B1} in post stimulus period averaged across stimulus intensities. **J)** Average distance to I_{B1} across stimulus intensities and simulated causal perturbations in the post stimulus period. **K)** Proportion of population in bursting regime across stimulus intensities and simulated causal perturbations in the post stimulus period. **L)** Average inter-compartment coupling probability across stimulus intensities and simulated causal perturbations in the post stimulus period. **M)** Average spike count across stimulus intensities and simulated causal perturbations in the post stimulus period.

ulus onset (Figure 3.2I-J). This increased the proportion of time each somatic compartment spent in the bursting regime (Figure 3.2K), which in turn increased the average spike count of $L5_{ET}$ cells across stimulus intensities, resulting in a reduction of the model's perceptual threshold (Figure 3.2F & M). Conversely, inhibition of the apical dendrites (pink) and the thalamus (orange) both resulted in an increase in the perceptual threshold (Figure 3.2G-H). Importantly, however, the mechanisms underlying the increase in the perceptual threshold differed across apical dendrite and thalamic inhibitory perturbations. Inhibition of the apical dendrites increased the average distance to bifurcation at I_{B1} in the apical compartment across the network (Figure 3.2I-J). In contrast, inhibition of the thalamus resulted in a comparatively minor reduction in the distance to bifurcation in the apical dendrites (Figure 3.2I-J)

but reduced the thalamus mediated inter-compartment coupling probability (Figure 3.2L), thereby reducing the probability that a back propagating action potential could reach the apical compartment, and the probability with which a plateau potential could switch the regime of the somatic compartment from regular spiking to bursting. Together, this resulted in a similar reduction in the proportion of cells in the bursting regime for both thalamic and apical dendrite inhibition, and likewise, a similar reduction in the average stimulus-evoked spike count, explaining the comparable increase in perceptual thresholds (Figure 3.2M).

3.2.3 Thalamocortical spiking model generalises to visual rivalry

We next sought to generalise our thalamocortical model of perceptual awareness to visual rivalry formalising and interrogating the hypothesis that the role played by pulvinar - $L5_{ET}$ loops in visual cortex is analogous to the role played by POm - $L5_{ET}$ loops in barrel cortex. To simulate visual rivalry we drove the model with input representing orthogonal gratings presented to each eye, typical of standard binocular rivalry experiments [26, 56], targeting the soma (i.e., basal dendrites) of $L5_{ET}$ cells on opposite sides of the ring with orthogonal orientation preferences (Figure 3.3A).

Due to the fact that inhibitory connectivity is broader than excitatory connectivity, delivering external drive to opposite sides of the ring shifts the model into a winner-take-all regime with burst-dependent persistent states on either side of the ring competing to inhibit one another. Importantly, through the accumulation of slow a hyperpolarising adaptation current in the somatic compartment of each $L5_{ET}$ cell (representing slow Ca^{2+} mediated K^{+} currents; [57, 58]), burst-dependent persistent states are only transiently stable leading to stochastic switches between persistent states characteristic of binocular rivalry (Figure 3.3B-C). We operationalised perceptual dominance (i.e., awareness of a percept at the exclusion of the other) in terms of the difference in average firing rate between the persistent bumps' of $L5_{ET}$ cell activity on opposite sides of the ring. In line with common practice (e.g. [59]) in models of rivalry, a population was counted as dominant when it had a firing rate 5 Hz $>$ than its competitor and lasted for longer than 250 ms (results were robust across a large range of threshold values). The initial competition for perceptual dominance was determined by which population of $L5_{ET}$ cells first established a recurrent loop with matrix-thalamus. This interaction allowed the recurrently connected population of cells to enter a bursting regime, at which point the population had sufficient activity to maintain a persistent state and inhibit the competing population into silence. In the previous chapter, we argued that this emergent property of corticothalamic loops is a good candidate for a cellular-level correlate of perceptual awareness [16, 20].

Crucially, our model provides a cellular-level explanation of spontaneous rivalry-induced perceptual switches in terms of these same mechanisms. As the reliable initiation of Ca^{2+} plateau potentials in the model depends upon back-propagating action potentials, perceptual switches are always preceded by regular spikes in the suppressed population. Preceding each switch, the accumulation of adaptation reduces the firing rate of the dominant population to a sufficient level that the suppressed population can escape from inhibition and emit a brief series of regular spikes that transition into bursts as thalamus is recruited and Ca^{2+} plateau potentials are initiated (Figure 3.3C), eventually gaining sufficient excitatory activity to inhibit the previously dominant population into silence via recurrent interactions with the basket cell population. At the population level (Figure 3.3D), this cascade of neuronal events is characterised by an initial ramping in the somatic firing rate of the suppressed population, followed

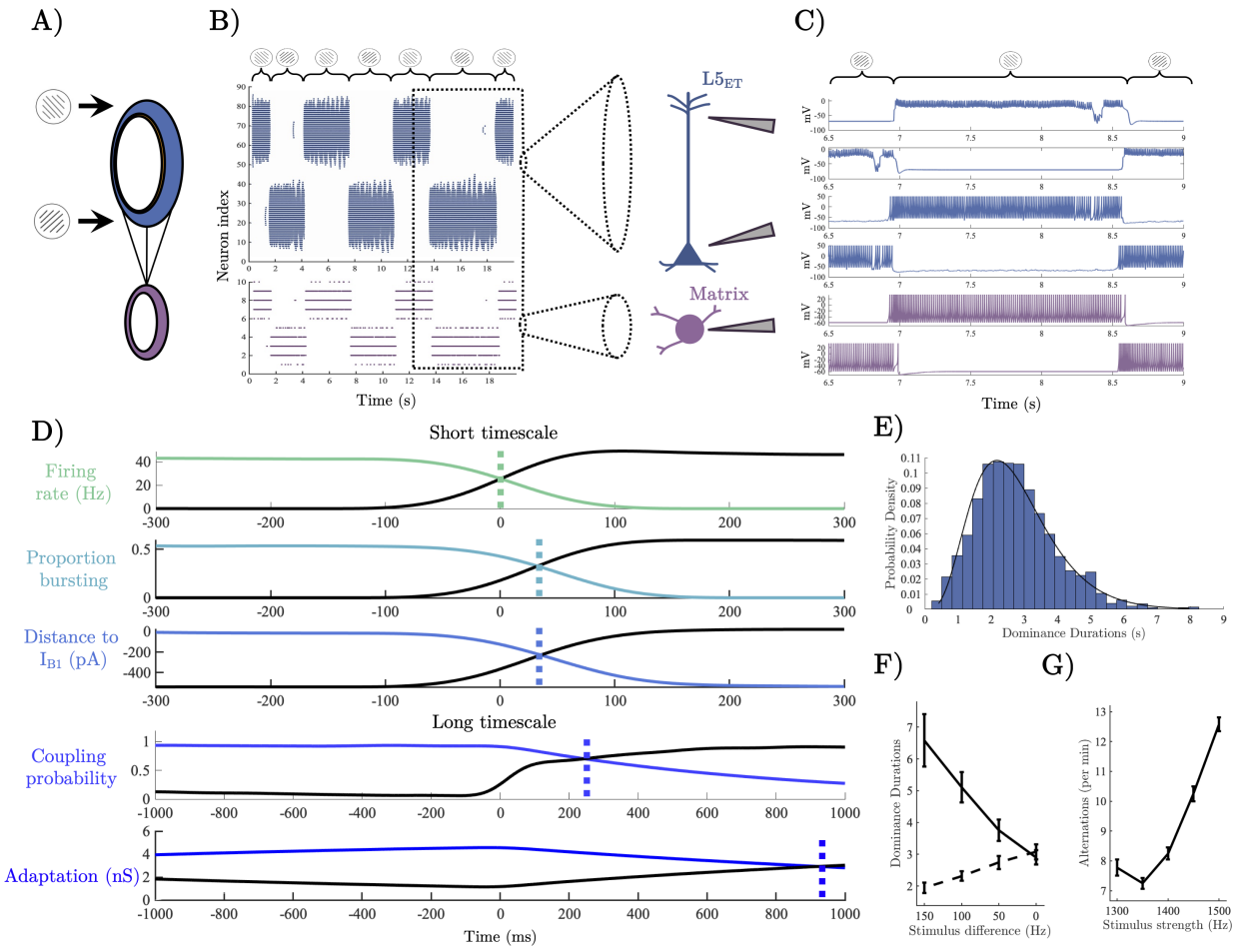

Figure 3.3: **A)** Thalamocortical ring architecture driven by input representing orthogonal gratings. **B)** Raster plots of $L5_{ET}$ soma (blue) and matrix thalamus population (purple) during rivalry. **C)** Example single neuron spiking activity around a perceptual switch - dotted lines denote the crossing in the population averaged time series. **D)** Population averaged neuronal variables centred on a perceptual switch. **E)** Histogram of dominance durations, black line shows the fit of a Gamma distribution with parameters estimated via MLE ($\alpha = 4.85$, $\theta = 0.56$). **F)** Simulation confirming Levelt's second proposition. Dashed line shows the dominance duration of the population receiving the decreasing external drive, solid line shows dominance duration of population receiving a fixed drive. **G)** Simulation of Levelt's fourth proposition. Notably, for low stimulus strengths our model, along with a number of mean-field models of visual rivalry [33], predicts a deviation from Levelt's fourth proposition.

by the inter-compartment coupling probability, the proportion of the population bursting, and finally by the average distance to bifurcation in the apical compartment. Once dominant, approximately half the population is in a bursting regime at each point in time (0.5424 mean $\pm$ 0.058 standard deviation), and the average distance to bifurcation in the apical compartment fluctuates around zero (-11.0185 $\pm$ 48.067 pA), with approximately a third (0.3669 $\pm$ 0.097) of the population located above the critical boundary (I_{B1}) at each point in time. In addition, the average inter-compartment coupling probability is (0.8845 $\pm$ 0.102) allowing reliable communication between compartments. Once silenced, adaptation in the previously dominant population decays back to baseline levels and accumulates in the previously suppressed, now dominant, population. In this way, the competitive organisation of the cortico-cortical and corticothalamic loop provides a natural flip-flop switch that stochastically alternates between dominant percepts.

As cortical pyramidal cells are known to display considerable differences in spiking behaviour across species [60] we swept the parameters of our model $L5_{ET}$ cell responsible for the bursting behaviour, and the inter-compartment coupling probability, to ensure that visual rivalry (quantified by average dominance duration) is stable across a wide range of parameters. In favour of the robustness of the model, as long as there was reliable inter-compartment coupling, which we assume is the case in the waking state, dominance durations in an empirically plausible range (i.e. with a period on the order of seconds) were present across a wide range of parameter values (supplementary material 3.5.3).

In close agreement with psychophysical data, switches between dominance and suppression had a right-skewed distribution of dominance durations (3.3E; [56, 61, 62]). Across stimulus drive conditions (1300 - 1500 Hz), comparison of negative log likelihoods showed that the distribution of dominance durations was best fit by a Gamma distribution ($l=2.81\times10^{-3}$), compared to lognormal ($l=2.88\times10^{-3}$) or normal distributions ($l=2.88\times10^{-3}$). Because of the large size of the network and the discontinuity in the somatic compartment we could not use analytic methods to interrogate the structure of the dynamical system underlying the stochastic oscillations. Instead, we used a heuristic line of argument (see [63], Ch.8) combined with simulations in the absence of noise to confirm that the dynamical regime underlying these stochastic oscillations likely consists of noisy excursions around a stable closed orbit (i.e. a stable limit cycle; 3.5.4).

3.2.4 Thalamocortical spiking model conforms to Levelt's propositions

To further test the psychophysical validity of our neurobiologically detailed model, we simulated the experimental conditions described by Levelt's modified propositions [61] - a set of four statements that compactly summarise the relationship between stimulus strength (e.g., luminance contrast, colour contrast, stimulus motion) and the average dominance duration of each stimulus percept. Here, we focus on the modified second and fourth propositions, as they constitute the core laws of rivalry and incorporate recent psychophysical findings (propositions one and three are consequences of proposition two; [61]).

Levelt's modified second proposition states that increasing the difference in stimulus strength between the two eyes will principally increase the average dominance duration of the percept associated with the stronger stimulus [56, 61, 64]. To simulate Levelt's second proposition, we decreased the spike rate entering one side of the ring from 1400 to 1250 Hz in steps of 50 Hz across simulations. In line with predictions (Figure 3.3F), the average dominance duration of the percept corresponding to the stronger stimulus showed a steep increase from ∼3 s with matched input, to ∼6.5 s with maximally different input while the average dominance duration on the side of the weakened stimulus decreased comparatively gradually to ∼2 s with maximally different inputs.

According to Levelt's modified fourth proposition, increasing the strength of the stimulus delivered to both eyes will increase the average perceptual reversal rate (i.e., decrease dominance durations; [61]), a finding that has been replicated across a wide array of experimental settings [65–68]. To simulate Levelt's fourth proposition, we ran a series of simulations in which we increased the spike rate of the external drive in steps from 1300 to 1500 Hz in steps of 50 Hz across simulations. Again in line with predictions (Figure 3.3G), the perceptual alternation rate increased with input strength, starting at ∼7.75 alternations per minute at the second weakest stimulus strength (1350 Hz) and increasing to ∼15 alternations per minute

for the strongest stimulus (1500 Hz). Interestingly, along with a number of mean-field models of rivalry [33], our model predicts a deviation from Levelt's fourth proposition for very low stimulus values with an uptick in alternation rate occurring at the lowest external dive value (1300 Hz). Encouragingly, there is some initial evidence that deviations from Levelt's fourth law may be present in human psychophysical data [61].

To help ensure that the simulation results were not biased by finite size effects or other simplifying assumptions such as the all-to-all connectivity of the cortical ring, or the 50/50 excitatory/inhibitory neuron ratio, we show in supplementary material 3.5.5 that a scaled-up version of the model consisting of 2000 cortical neurons with sparse connectivity, and an 80/20 excitatory/inhibitory neuron ratio (i.e. consistent with Dale's law), also produces a Gamma distribution of dominance durations, and is consistent with Levelt's second and fourth propositions. We thus confirmed that our neurobiologically detailed model of the matrix thalamus - $L5_{ET}$ loop is capable of reproducing Levelt's propositions, which together with the right-skewed distribution of dominance durations, show the consistency of our model with the psychophysical laws known to govern visual rivalry.

3.2.5 Generating testable predictions through *in silico* electrophysiology

Binocular rivalry is thought to depend in part on the substantial degree of binocular overlap in humans ($\sim$120°), however the lateral position of the eyes in mice leaves only $\sim$40° of binocular overlap [69]. For this reason, there are no current mouse models of binocular rivalry, however there are monocular variants of visual rivalry, namely plaid perception, that can be studied in mouse models [13, 28]. Crucially, plaid perception, like binocular rivalry, conforms to Levelt's laws [27, 61] and also has a right skewed distribution of dominance durations that is well fit by a Gamma distribution [28]. We hypothesise, therefore, that the principles underlying the simulation of binocular rivalry in our model will also describe other forms of visual rivalry (such as plaid perception), offering a plausible means to test cellular level predictions derived through simulation. To this end, we next ran a series of perturbation experiments, with the aim of interrogating the novel burst-dependent mechanism of perceptual dominance by mimicking the optogenetic and pharmacological experiments carried out in threshold-detection studies [18, 19], in the context visual rivalry.

As perceptual dominance depends on the formation and maintenance of a burst-dependent persistent state, we hypothesised that artificially exciting the apical dendrites would result in an increase in the average dominance duration of the excited population, and artificially inhibiting the apical dendrites and thalamus would result in a decrease in the dominance durations for the inhibited populations. Due to the fact that the distance to bifurcation of the dominant population fluctuates around zero, we predicted that exciting the apical dendrites would increase the proportion of the population above the bifurcation at I_{B1}, thereby reducing the probability with which a fluctuation in somatic drive would lead to a sizable enough drop in the proportion of the population below I_{B1} to release the competing population from inhibition. This should, therefore, result in an increase in the frequency of long dominance duration events, thereby increasing the mean and the spread of the distribution. Equivalently, we predicted that inhibiting the apical dendrites would reduce the proportion of the population above I_{B1}, making it more likely that transient fluctuations in somatic drive would allow the competing population to escape from inhibition, reducing the occurrence of long dominance duration events, thereby reducing both the mean and the spread of the distribution of

dominance durations. Finally, based on the results of the threshold detection simulations we predicted that thalamic inhibition would have an analogous effect on dominance durations to apical dendrite inhibition but would be mediated by a reduction in the coupling probability.

To test these hypotheses, we conducted two *in silico* experiments analogous to the conditions described by Levelt's modified propositions but instead of manipulating the external drive entering the somatic compartment of $L5_{ET}$ cells we manipulated the amplitude of simulated causal perturbations to the $L5_{ET}$ apical compartment and thalamus (Figure 3.4A & 3.5A).

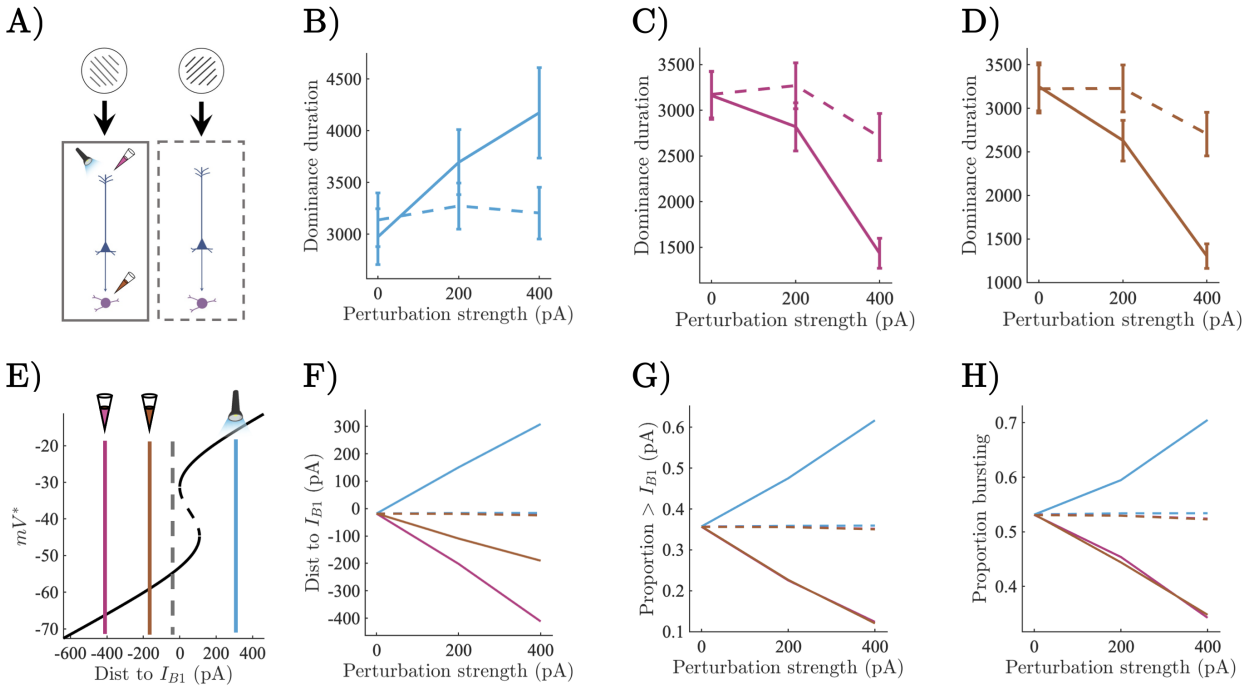

Figure 3.4: **A)** Causal perturbations to one half of the thalamocortical circuit underlying visual rivalry consisting of optogenetic excitation of the apical compartment (blue), pharmacological inhibition of the apical compartment (pink), and pharmacological inhibition of the thalamus (orange). **B-D)** Average dominance duration of perturbed (solid), and unperturbed (dashed) populations. Error bars show SEM. **E)** Average distance to bifurcation point at I_{B1} shown on bifurcation diagram for perturbed (solid) and unperturbed (dashed) populations during periods of perceptual dominance with 400 pA perturbation strength. **F)** Average distance to bifurcation point at I_{B1} for perturbed (solid) and unperturbed (dashed) populations during periods of perceptual dominance. **G)** Proportion of population above bifurcation point at I_{B1} during periods of perceptual dominance. **H)** Proportion of population in bursting regime during periods of perceptual dominance.

In the first set of experiments, we simulated optogenetic excitation and pharmacological inhibition of one of the two competing populations by adding a constant current (± 200, 400 pA) to all of the target variables (i.e., apical compartment or thalamic neurons) on one side of the ring (Figure 3.4A). In line with predictions, we found that the average dominance duration of the excited population (Figure 3.4B) increased, the distance to bifurcation decreased (Figure 3.4E-F), the proportion of the population above the critical point at I_{B1} increased (Figure 3.4G), and the proportion of the population in the bursting regime increased (Figure 3.4H). The dominance durations and neuronal dynamics of the unexcited population remained relatively unchanged. Similarly, inhibition of both the apical dendrites and thalamus reduced the average dominance duration of the inhibited population while the uninhibited population was again relatively unchanged (Figure 3.4C-D). As in the threshold detection simulations, inhibition of the apical dendrites led to a large increase in the distance to I_{B1} compared to thalamic inhibition (Figure 3.4E-F) which primarily affected the inter-compartment coupling

probability (supplementary material 3.5.6). Both apical dendrite and thalamic inhibition led to almost identical reductions in the proportion of the population above the critical point at I_{B1} (Figure 3.4G), and the proportion of the population in the bursting regime (Figure 3.4H). The uninhibited population again remained relatively constant across all of the neuronal measures (the small drop in dominance durations of the uninhibited population for the 400 pA inhibitory perturbations is due to the adaptation variable having less time to recover). As predicted the spread of the distribution of dominance durations increased with the amplitude of excitatory perturbation and decreased under inhibition (supplementary material 3.5.6).

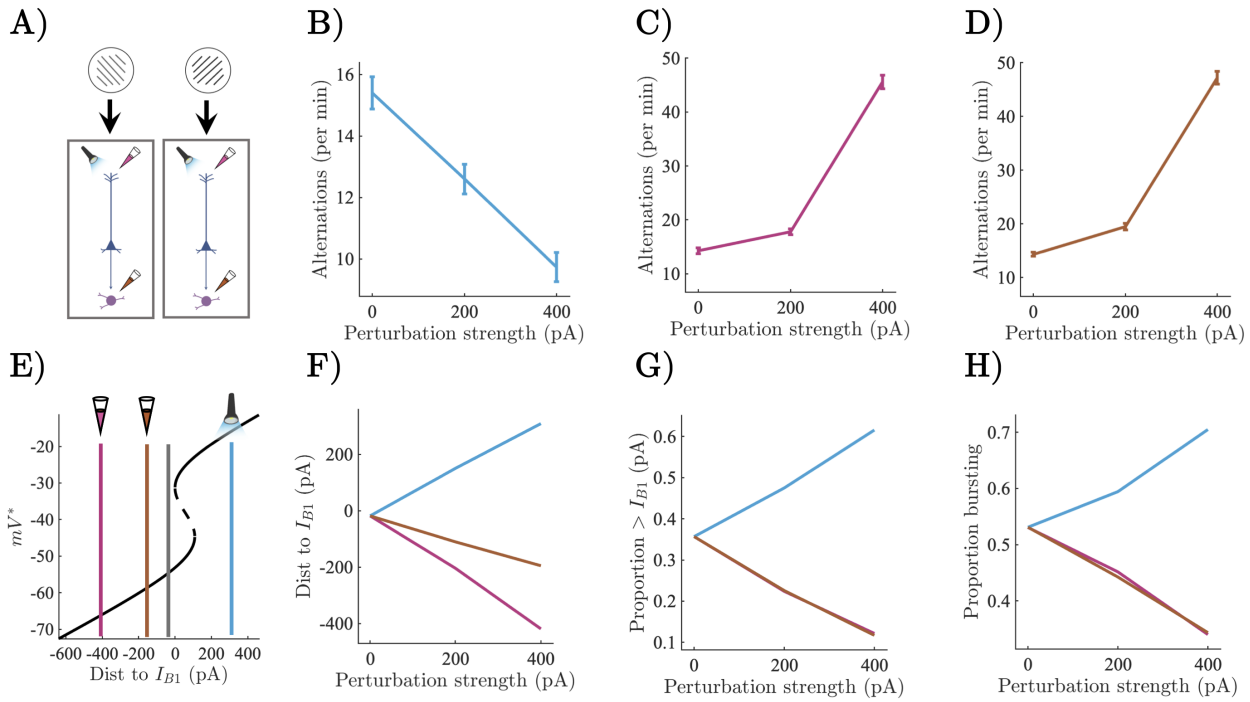

Figure 3.5: **A)** Causal perturbations to full thalamocortical circuit underlying visual rivalry. Colours same as above. **B-D)** Perceptual alternations per minute of simulation time across perturbation types. Error bars show SEM. **E)** Average distance to bifurcation point at I_{B1} shown on bifurcation diagram for perturbed (blue, orange, pink) and unperturbed (grey) simulations during periods of perceptual dominance at 400 pA. **F)** Average distance to bifurcation point at I_{B1} for full network perturbations during periods of perceptual dominance. **G)** Average proportion of population above bifurcation point at I_{B1} for full network perturbations during periods of perceptual dominance. **H)** Average proportion of population in bursting regime for full network perturbations during periods of perceptual dominance.

In the second set of experiments, we simulated optogenetic excitation and pharmacological inhibition of both competing neuronal populations simultaneously by adding a constant current (± 200, 400 pA) to all of the target variables on the ring (Figure 3.5A). Again, in line with predictions, the speed of rivalry (i.e., the number of perceptual alternations per minute of simulation time) decreased as a function of apical dendrite excitation (Figure 3.5B). Excitation also decreased the average distance to I_{B1} (Figure 3.5E-F), increased the proportion of the population above the critical point at I_{B1} (Figure 3.5G), and increased the proportion of the population in the bursting regime (Figure 3.5H). In contrast, inhibition of the apical dendrites and thalamus both increased the speed of rivalry (Figure 3.5C-D). Inhibition of the apical dendrites led to a large increase in the distance to the critical point at I_{B1} compared to thalamic inhibition (Figure 3.5E-F), but both apical dendrite inhibition and thalamic inhibition reduced the proportion of the population above the critical point at I_{B1} (Figure 3.5G), and

the proportion of the population in a bursting regime (Figure 3.5H) through reductions in the thalamus mediated inter-compartment coupling probability (supplementary material 3.5.6). As with the asymmetric perturbation simulations, inhibition exerted a much larger effect on the speed of rivalry than excitation. Finally, again in line with our predictions, the spread of the distribution of dominance durations increased under excitatory perturbation of the apical dendrites and decreased under inhibition of the apical dendrites and thalamus (supplementary material 3.5.6). Together, these *in silico* electrophysiological experiments provide important testable (and explainable) hypotheses for future experiments that although not testable in any existing data sets are well within the purview of modern systems neuroscience providing an opportunity to conduct precise theory-driven tests of the model.

3.3 Chapter 3 discussion

The study of perceptual awareness in human participants and animal models has so far proceeded largely in parallel - the former exploring the largescale neural dynamics and behavioural signatures of perceptual awareness across a rich array of experimental settings, and the latter characterising the cellular circuitry of perception in exquisite detail, and with precise causal control, but with only limited links to higher level perceptual phenomena [14]. Leveraging a neurobiologically detailed model of the matrix thalamus - $L5_{ET}$ loop, we have shown that a potential circuit-level mechanism of perceptual awareness discovered in a mouse model of tactile awareness [15, 18, 19] generalises to visual rivalry, thus providing a roadmap for linking the circuit level mechanisms studied in animal models to the behavioural signatures of perceptual awareness studied in human psychophysics.

The balance of neurobiological detail and interpretability offered by our model allowed us to reproduce the threshold-detection results of [18, 19] and interrogate the mechanisms underlying the experiments in a manner that would be impossible *in vivo*. In particular, examination of the model's dynamics under simulated causal perturbations to the circuit revealed a degenerate dynamical mechanism for controlling the threshold for perceptual awareness. Excitation of the apical compartment reduced the distance to bifurcation in the apical compartment, thus increasing the probability that each cell would generate a Ca^{2+} plateau potential, switching the soma into a bursting regime. This resulted in an increase in the baseline and stimulus-evoked spike count, and correspondingly, led to a reduction in the model's perceptual threshold. Inhibition of the apical compartment and thalamus resulted in comparable downward shifts in the baseline and stimulus-evoked spike count, leading to increases in the model's perceptual threshold. Importantly, the neural mechanisms underlying the increases in perceptual threshold were distinct: inhibiting the apical compartment increased the distance to bifurcation, thus reducing the probability with which each cell could generate a Ca^{2+} plateau potential, whereas inhibiting the thalamus reduced the inter-compartment coupling. Both mechanisms, however, led to comparable reductions in the proportion of cells in the bursting regime explaining the comparable increase in perceptual thresholds, suggesting that it is the emergent action of the corticothalamic circuit as a whole, rather than single cells within the circuit, that are responsible for perceptual awareness.

The degenerate mechanisms underlying the threshold for perceptual awareness combined with the operational definition of perceptual awareness in the threshold detection task (in terms of psychometric functions) points to a conceptually important point about the role of bursts in the model, and potentially, the empirical data itself. Specifically, controlling the ease with which a cell can burst through optogenetic and pharmacological perturbation is simply a

means for controlling how easily a stimulus can evoke reverberant activity in corticocortical and thalamocortical loops which, in the simple case of threshold detection, constrains the extent to which stimulus evoked activity can stand out against a background of noise driven fluctuations.

We next showed that the same thalamus-gated burst-dependent mechanism underlying perceptual awareness in simulations of the tactile threshold detection task also determines perceptual dominance in simulations of visual rivalry. Specifically, perceptual dominance is initiated by a succession of regular spikes and maintained through the formation of a transiently stable burst-dependent persistent state characterised by reliable coupling between apical and somatic compartments. This allows the apical compartment to generate temporally extended plateau potentials in a large subset of the dominant population, reliably switching the $L5_{ET}$ soma from a regular spiking to a bursting regime. Perceptual dominance is then maintained until the slow hyperpolarising adaptation current accumulates to a sufficiently high level that the dominant population is no longer able to inhibit the competing population into silence, and a perceptual switch ensues. Importantly, the model conforms to Levelt's modified propositions. Originally proposed in 1965 [70], Levelt's laws have proven to be remarkably robust needing only minor modification and contextualisation [61] and have, therefore, served as a benchmark for computational models of visual rivalry (e.g., [29, 31, 33, 36]). Together with the right-skewed (Gamma) distribution of dominance durations the consistency of our model with Levelt's propositions provides an *in silico* conformation of the hypothesis that pulvinar - $L5_{ET}$ loops in visual cortex may play an analogous role to POm - $L5_{ET}$ loops in barrel cortex. This is a minimal but necessary first step in testing the hypothesis that reverberant activity in matrix thalamus - $L5_{ET}$ loops is a necessary component part in a domain-general mechanism of perceptual awareness.

Having validated our model against psychophysical benchmarks, we next sought to interrogate the novel thalamus-gated burst-dependent mechanism of perceptual dominance by emulating the optogenetic and pharmacological experiments carried out by [18, 19] in the context of visual rivalry. Under conditions of visual rivalry, the simulated causal perturbations are similar to the conditions described by Levelt's propositions, but instead of manipulating the strength of the external stimulus we manipulated the strength apical compartment excitation/inhibition, or thalamic inhibition, highlighting the unique contribution of these neurobiological components to visual rivalry. Across asymmetric and symmetric perturbations excitation of the apical compartment slowed perceptual alternations (i.e., increased dominance durations) by increasing the proportion of the population able to sustain temporally extended Ca^{2+} plateau potentials and remain in a transiently-stable bursting regime, whereas inhibition of both the apical dendrites and thalamus had the opposite effect. Although technically difficult, these simulated experimental manipulations are well within the purview of modern experimental techniques and therefore represent a means of causally testing the predictions of our model. Importantly, the simulation of these experimental perturbations would not be possible in any existing models of rivalry, even those at the spiking level (e.g. [31, 35, 50]), as they focus on the minimal conditions for rivalry in point-neuron models of cortical interaction. The inclusion of a dual compartment model of $L5_{ET}$ cells, and an explicit thalamic population, was, therefore, required in order to make contact with the results of [18, 19]. Indeed, although (thalamus-gated) plateau potential induced bursting is necessary for rivalry in our model, if we were to coarse-grain the model the dynamics would be well described by a mean-field model tracking only the mean firing rate of each population. A line of reasoning we pursue in part two of this thesis. Thalamic control of $L5_{ET}$ bursting is therefore not strictly speaking

necessary for modelling the mesoscale dynamics of binocular rivalry, but it is necessary if one wants to capture the neurobiological processes that underpin the mesoscale dynamics and how cellular-level interventions such as optogenetic excitation or pharmacological inhibition impact rivalry dynamics as we have done here.

In addition to the predicted effect of causal perturbations on visual rivalry, our model generates a number of more straightforward correlational predictions. Specifically, matrix-rich higher-order thalamic nuclei with recurrent connections to sensory cortex, such as the pulvinar, should be selective for perceptual awareness rather than physical stimulation, a prediction supported by both human neuroimaging [3, 4] and non-human primate electrophysiology [71]. Similarly, synchronous bursting activity in deep layers of cortex, specifically $L5_B$, which contain the soma of $L5_{ET}$ cells, should likewise be selective for perceptual awareness rather than physical stimulation, a prediction that, with the advent of primate Neuropixels [72], is also readily testable. Finally, in the context of visual rivalry, perceptual dominance should be characterised by elongated Ca^{2+} plateau potentials in the apical dendrites of $L5_{ET}$ cells (located in L1) in cells selective for the dominant percept, a prediction testable in mouse models of visual rivalry (e.g. [28]).

We anticipate that the cellular conditions for awareness explored in this paper are likely to have consequences for the large-scale correlates of awareness. Indeed, we venture that at the level of large-scale brain networks diffuse matrix-thalamus gated bursting may play a key role in the formation of a quasi-critical regime [73, 74] allowing single nodes in a network to transiently escape from a tight E/I balanced state. This permits stimulus information to rapidly propagate across the cortical sheet while also maintaining stability at the level of the whole network [73], effectively modulating the gain of interareal connectivity in line with previous computational models of pulvinar-cortical interactions in cognitive tasks [75]. Indeed, efforts to test the large-scale consequences of the cellular level mechanisms interrogated in this paper are already underway. Biophysical modelling of source-localised MEG data showed that auditory awareness evoked activity is best fit by increased input to superficial layers of the cortical column, consistent with the projections of matrix-type higher-order thalamus [76].

To strike the right balance between neurobiological detail and interpretability, we made a number of simplifying assumptions that place some limitations on our model. Most notably, we did not include L2/3 pyramidal neurons - which are arguably the primary source of long distance horizontal connections in the cortex [77] and arguably cross column inhibition [3] - nor a core thalamic population which forms a targeted recurrent loops with L4 and L6 of cortex [78] preventing us from performing systematic perturbation experiments on our model highlighting the precise function of $L5_{ET}$ cells and higher-order matrix thalamus in a more realistic cortical microcircuit. We also did not include time delays between our corticocortical or thalamocortical connections, preventing us from modelling time-frequency components of common electrophysiological measures such as local field potentials (e.g., [79]). Finally, our model has only a single hierarchical level, preventing us from making contact with evidence showing a potential prefrontal contribution to perceptual switches [80, 81].

Our model is, of course, only a first step towards a formal characterisation of the minimal neurobiological mechanisms underlying perceptual awareness. Extending the model, and modelling strategy more generally, to new paradigms such as backward masking [11], will be of paramount importance in the progression of the field as mouse models and the tools of systems neuroscience are brought into contact with the sophisticated psychophysical paradigms used to study the behavioural signatures of awareness in humans.

3.4 Chapter 3 methods

3.4.1 Thalamocortical spiking neural network

The neuronal backbone of the model consists of a (novel) dual compartment model of $L5_{ET}$ neurons, fast spiking interneurons (basket cells), and thalamic cells. The dynamics of basket cells, thalamic cells, and the somatic compartment of $L5_{ET}$ cells were described by Izhikevich quadratic adaptive integrate and fire neurons, a hybrid dynamical system that is capable of reproducing a wide variety of spiking behaviour while still being highly efficient to integrate numerically [38–40]. The Izhikevich neuron consists of the following two-dimensional system of ODEs:

$$\begin{aligned} C\dot{v}^{(s)} &= k(v^{(s)} - v_r^{(s)})(v^{(s)} - v_t^{(s)}) - u^{(s)} + I_{ext} \\ \dot{u}^{(s)} &= a(b(v^{(s)} - v_r^{(s)}) - u^{(s)}), \end{aligned} \tag{3.1}$$

with reset conditions: if $v \geq v_{peak}$ then $v \rightarrow c$, $u \rightarrow u + d$. The equations are in dimensional form giving the membrane potential (including the resting potential v_r, spike threshold v_t, and spike peak v_{peak}, and reset c), input I_{ext}, time t, and capacitance C, biophysically interpretable units (mV, pA, mS, and pF respectively). The remaining four parameters k, a, b, and d, are dimensionless and control the sharpness of the quadratic-nonlinearity, the timescale of spike adaptation, the sensitivity of spike adaptation to sub-threshold oscillations, and the magnitude of the spike reset adaptation variable. Crucially, Izhikevich [40, 82] fit parameters for a large class of cortical and sub-cortical neurons, thus affording our model a high degree of neurobiological plausibility while greatly reducing the number of free parameters.

The apical compartment of the $L5_{ET}$ neuron consists of a two dimensional non-linear system introduced by Naud and colleagues [41] as a phenomenological model of the Ca^{2+} plateau potential in the apical dendrites of $L5_{ET}$ neurons,

$$\begin{aligned} C\dot{v}^{(d)} &= -l(v^{(d)} - v_r^{(d)}) + gf(v^{(d)}) + mH(t - t_s) + u^{(d)} + I_{ext} \\ \dot{u}^{(d)} &= a(b(v^{(d)} - v_r^{(d)}) - u^{(d)}). \end{aligned} \tag{3.2}$$

Where $f(x) = 1/(1+\exp(-\frac{(x+38)}{6})$ describes the regenerative non-linearity underlying the Ca^{2+} plateau potential, and $H(t-t_s)$ denotes a square wave function of unitary amplitude describing the backpropagating action potential (delayed by 0.5 ms and lasts for 2 ms) with t_s denoting the somatic compartment spike time. The parameters l, g, m, denote the leak conductance nS, amplitude of regenerative non-linearity pA, and amplitude of the back propagating action potentials pA respectively. The model and parameters were derived from a more complex model of Ca^{2+} spikes built to predict *in vitro* $L5_{ET}$ spike times [47].

To simulate key observations from empirical experiments, we coupled the compartments together so that sodium spikes in the somatic compartment triggered a back-propagating action potential affecting the apical compartment through the square wave function $H(t - t_s)$. In turn, plateau potentials in the apical compartment controlled the reset conditions of the somatic compartment. We leveraged the insight [38, 45, 46] that the difference between regular spiking and intrinsic bursting can be modelled by changing the reset conditions of equation (3.1), raising the reset voltage (increasing c) taking the neuron closer to threshold, and reducing the magnitude of spike adaptation (decreasing d). Whenever the membrane potential in

the apical compartment exceeded -30 mv the reset conditions changed from regular spiking to bursting parameters. This allowed us to reproduce the transient change in dynamical regime in $L5_{ET}$ cells that occurs when they receive coincident apical and basal drive. Parameters values for each neuron/compartment are given in Table 3.4.1.

Neuron	C pF	k a.u.	v_r mV	v_t mV	a	b	c mV	d a.u.	v_{peak} mV	l nS	g pA	m pA
$L5_{ET}$ apical dendrite	170	∼	-70	∼	130 ms $^{-1}$	-13 nS	∼	∼	∼	24.2857	1200	2600
$L5_{ET}$ soma	150	2.5	-75	-45	0.01 a.u	5 a.u	RS: -65, IB: -55	RS: 250, IB: 150	50	∼	∼	∼
Basket cell	20	1	-55	-40	0.15 a.u	8 a.u	-55	200	25	∼	∼	∼
Matrix thalamus	200	1.6	-60	-50	0.01 a.u	15 a.u	-60	10	35	∼	∼	∼

Table 3.1: Parameters for each neuron. $L5_{ET}$ apical dendrite parameters were taken from [41]. L_{ET} soma parameters were modified from the model of intrinsic bursting (p.290) described in [40]. Basket cell (fast spiking interneuron), and matrix thalamus parameters were taken from [82].

Based on the finding that communication between apical dendrites and the soma of $L5_{ET}$ cells requires depolarising input from the matrix thalamus to the apical coupling zone in L5a [48] we made back propagating action potentials and Ca^{2+} driven parameter switches depend stochastically upon a phenomenological model of excitatory dynamics in the apical coupling zone described by the following saturating linear system:

$$\dot{g}_{i,\text{coupling}} = -\frac{g_{i,\text{coupling}}}{\tau_{\text{coupling}}} + (1 - g_{i,\text{coupling}}) \sum_j \delta(t - t_j^s). \tag{3.3}$$

Coupling was driven by thalamic spikes (where t_s denotes the time that the thalamic neuron passes the threshold $v \geq v_{peak}$) and the decay constant $\tau_{coupling}$ was taken from work estimating the decay of the post synaptic excitatory effects of metabotropic glutamate receptors [83] which have been shown empirically to mediate inter-compartmental coupling in $L5_{ET}$ cells [48]. By design, the dynamics of the coupling variable varied between 0 and 1 and governed the probability with which back-propagating action potentials would reach the apical compartment and the probability with which a Ca^{2+} spike would lead to a switch in the soma reset parameters.

Based on previous spiking neural network models of rivalry [30, 31, 50], the cortical component of the network had a one-dimensional ring architecture. Each point on the ring represents an orientation preference with one full rotation around the ring corresponding to a 180° visual rotation. This mirrors the fact that a 180° rotation of a grating results in a visually identical stimulus and also ensures periodic boundary conditions. The cortical ring contained 90 $L5_{ET}$ neurons and 90 fast spiking interneurons. Each pair of excitatory and inhibitory neurons was assigned to an equidistant location on the ring (unit circle) giving each neuron a 2° difference in orientation preference relative to each of its neighbours. The (dimensionless) synaptic weights $w_{i,j}^{\kappa\omega}$ connecting neurons (E→E, I→E, and E→I), were all-to-all with amplitude decaying as a function of the Euclidean distance $d_{i,j}$ between neurons according to a spatial Gaussian footprint:

$$d_{i,j} = \sqrt{(\cos\theta_i - \cos\theta_j)^2 + (\sin\theta_i - \sin\theta_j)^2}, \tag{3.4}$$

$$w_{i,j}^{\kappa\omega} = \lambda^{\kappa\omega} e^{-\frac{1}{2}(\frac{d_{i,j}}{\sigma^{\kappa\omega}})^2}. \tag{3.5}$$

Where θ is the location of the neuron on the unit circle, κ and ω denote the pre- and post-synaptic neuron type (i.e. E→E), λ controls the magnitude of the synaptic weights, and $\sigma^{\kappa\omega}$ the spatial spread. In line with empirical constraints inhibitory coupling had a larger

spatial spread than excitatory to excitatory coupling [84]. Each thalamic neuron received input from 9 cortical neurons and then projected back up to the apical dendrites of the same 9 cortical neurons recapitulating the diffuse projections of higher-order thalamus onto the apical dendrites of $L5_{ET}$ neurons in layer 1 [52, 53]. For simplicity we set projections to and from the thalamus to a constant value (e.g., $w_{i,j}^{TH\to D} = \lambda^{TH\to D}$). For the sake of computational efficiency we also neglected differences in rise time between receptor types which allowed us to model receptor dynamics with a first-order linear differential equation with decay (τ_{decay}) constants chosen to recapitulate the dynamics of inhibitory ($GABA_A$), and excitatory (AMPA and NMDA) synapses [85, 86]:

$$\dot{g}_{i,\text{syn}} = -\frac{g_{i,\text{syn}}}{\tau_{\text{decay}}} + w_{i,j}^{\kappa\omega} \sum_j \delta(t - t_j^s). \tag{3.6}$$

Where, as above, t_s denotes the time that the neuron passes the threshold $v \geq v_{peak}$. The conductance term entered into the input I_{ext} through the relation $I_{i,syn} = g_\infty g_{i,syn}(E_{syn} - v_i(t))$ where E_{syn} is the reverse potential of the synapse. Following [82] we set g_∞ to 1 for $GABA_A$ and AMPA synapses, and $g_\infty = \frac{((v+80)/60)^2}{1+((v+80)/60)^2}$ for NMDA synapses. To prevent artificial distortions of the spike shape that can occur during parameter sweeps that push the model outside its normal operating regime we clipped individual NMDA conductances to a maximum value of 85 nS.

For the threshold detection simulations, the somatic compartment of each $L5_{ET}$ cell received 600 Hz of independent (Poisson) external drive and apical compartments received 50 Hz of external drive. The whisker deflection was simulated by a pulse of constant amplitude varying between 0 - 350 pA lasting 200 ms and weighted by the spatial Gaussian shown in equation (3.7) where N is the neuron at the centre of the pulse and σ_{TD} the spatial spread:

$$h_i^{TD} = e^{-(\frac{(i-N)}{\sigma^{TD}})^2}. \tag{3.7}$$

For the visual rivalry simulations, separate monocular inputs targeting the somatic compartment of $L5_{ET}$ cells were modelled with two independent Poisson processes (representing input from the left and right eyes in the case of binocular rivalry or left and right movement selective populations in the case of plaid perception) with rates varying between 1200 and 1800 Hz depending on the simulation. The external drive was weighted by the spatial Gaussian shown in equation (3.8) centred on neurons 90° apart on the ring abstractly corresponding to the orthogonal grating stimuli commonly employed in binocular rivalry experiments:

$$h_i^{VR} = e^{-(\frac{(i-N_L)}{\sigma^{VR}})^2} + e^{-(\frac{(i-N_R)}{\sigma^{VR}})^2}. \tag{3.8}$$

Here i denotes the index of the i-th neuron, N_L and N_R control the orientation of the stimulus delivered to the left and right eyes, and σ_{VR} the spatial spread. To capture the slow hyperpolarising current that traditionally governs switching dynamics in models of bistable perception [58], the somatic compartment of each $L5_{ET}$ cell was coupled to a phenomenological model of slow hyperpolarising Ca^{2+} mediated K^+ currents [57] which entered into the external drive term for each cell (i.e. $I_{i,adapt} = g_{i,adapt}(E_{adapt} - v_i(t))$) with dynamics given by:

$$\dot{g}_{\text{adapt}} = -\frac{g_t^{\text{adapt}}}{\tau_{\text{adapt}}} + \Delta g \sum \delta(t - t_s). \tag{3.9}$$

Where Δg denotes the contribution of each spike to the hyperpolarising current, and τ_{adapt} the decay constant.

Rather than fit the parameters of our model to individual experimental findings, which permits substantial degrees of freedom and risks over interpretation of idiosyncratic aspects of individual experiments, we instead elected to challenge a single model to qualitatively reproduce a wide array of experimental findings with a minimal set of parameter changes carefully chosen to reflect experimental manipulations and perturbations. Specifically, we initialised the connectivity parameters such that: 1) when the model received a background drive the conductances were approximately E/I balanced with a coefficient of variation > 1, corresponding to an asynchronous irregular regime [54]; 2) inhibitory connections on the cortical ring had broader (Gaussian) connectivity than excitatory connections generating a winner-take-all regime when the model received two competing inputs to opposite sides of the cortical ring; and 3) a slow hyperpolarising current was added to the somatic compartment of each $L5_{ET}$ cell destabilizing the winner-take-all attractor states leading to spontaneous switches between transiently stable persistent states with an average duration in the experimentally observed range for binocular rivalry. Parameters for the model components described by equations (3.3) - (3.9) are supplied in Table 3.4.1.

Parameter	**Description**	**Value**	**Units**
$\lambda^{E\to E}_{AMPA}$	Amplitude of excitatory to excitatory coupling for (AMPA)	$\frac{6.125}{\sigma^{E\to E}\sqrt{2\pi}}$	a.u.
$\lambda^{E\to E}_{NMDA}$	Amplitude of excitatory to excitatory coupling (NMDA)	$\frac{1.225}{\sigma^{E\to E}\sqrt{2\pi}}$	a.u.
$\lambda^{E\to I}$	Amplitude of excitatory to inhibitory coupling (NMDA and AMPA)	$\frac{1}{\sigma^{E\to I}\sqrt{2\pi}}$	a.u.
$\lambda^{I\to E}$	Amplitude of inhibitory to excitatory coupling ($GABA_A$)	$\frac{5}{\sigma^{I\to E}\sqrt{2\pi}}$	a.u.
$\lambda^{E\to TH}$	Constant excitatory to thalamic coupling constant (AMPA only)	4	a.u.
$\lambda^{TH\to D}_{AMPA}$	Constant thalamic to apical dendrite coupling (AMPA)	10	a.u.
$\lambda^{TH\to D}_{NMDA}$	Constant thalamic to apical dendrite coupling (NMDA)	10	a.u.
$\sigma^{E\to E}$	Spread of excitatory to excitatory coupling	0.5	a.u.
$\sigma^{E\to I}$	Spread of excitatory to inhibitory coupling	2	a.u.
$\sigma^{I\to E}$	Spread of inhibitory to excitatory coupling	2	a.u.
$\tau_{decay:AMPA}$	Decay time of AMPA conductance	6	ms
$\tau_{decay:GABA_A}$	Decay time of $GABA_A$ conductance	6	ms
$\tau_{decay:NMDA}$	Decay time of NMDA conductance	100	ms
$\tau_{coupling}$	Decay time of apical coupling zone	800	ms
τ_{adapt}	Decay time of adaptation current	2000	ms
Δg	Contribution of each spike to adaptation current	0.065	nS
σ_{TD}	Spatial spread of external drive	20	a.u.
σ_{VR}	Spatial spread of apical drive	18	a.u.
$E_{Excitatory}$	Reverse potential of excitatory synapses	0	mV
$E_{Inhibitory}$	Reverse potential of inhibitory synapses	-75	mV
E_{Adapt}	Reverse potential of adaptation currents	-80	mV

Table 3.2: Parameter description, values and units for the model components described by equations 3.1-3.9

The equations were integrated numerically in MATLAB 2023b. The apical compartment was integrated with a standard forward Euler scheme. All other compartments were integrated using the hybrid scheme for conductance-based models introduced by [87]. All simulations used a step size of 0.1 ms and were run for 30 s. Unless stated otherwise, all simulation results were averaged over a minimum of 30 random seeds.

3.4.2 Distance to bifurcation

To obtain a closed form expression for the distance to bifurcation in the apical compartment (equation 3.2), we leveraged the fact that saddle node bifurcations occur when the nullclines ($\dot{v}^{(d)} = \dot{u}^{(d)} = 0$) of the system intersect tangentially [63]. That is, the nullclines and derivative of the nullclines must be equal, leading to the following two requirements (where we have absorbed the term describing back propagating action potentials $H(t - t_s)$ into the external drive I_{ext}, which we treat as a constant):

$$\frac{l}{C}(v^{(d)} - v_r^{(d)}) + \frac{1}{C}(gf(v^{(d)}) + I_{ext}) = b(v^{(d)} - v_r^{(d)}), \tag{3.10}$$

$$\frac{d}{dv^{(d)}}\left[\frac{l}{C}(v^{(d)} - v_r^{(d)}) + \frac{1}{C}(gf(v^{(d)}) + I_{ext})\right] = \frac{d}{dv^{(d)}}\left[b(v^{(d)} - v_r^{(d)})\right]. \tag{3.11}$$

We used equation (3.11) to solve for $v^{(d)}$ giving $v^{(d)*} = -44.6601, -31.3399$. We then substituted $v^{(d)*}$ back into equation (3.10) to solve for I_{ext} yielding the value of the external current at each of the two bifurcations $\mathrm{I_{B1}} = 538.911$ pA and $\mathrm{I_{B2}} = 647.375$ pA (corresponding to points at which the linear adaptation current nullcline intersects tangentially with the left and right knees of the cubic membrane potential nullcline see supplementary material 3.5.1).

3.4.3 Psychometric and neurometric functions

To obtain a measure of response probability from our model comparable to the psychometric functions in [18, 19] we took a two-pronged approach. First, for all stimulus intensities including stimulus-absent trials (when the model only received a background drive) we calculated the frequency with which the spike count in the 1000 ms post-stimulus window exceeded a criterion defined on the interval between the minium and maximium spike count. We then selected the (optimal) criterion that best minimised misses and false alarms. Trials exceeding the optimal criterion were counted as a response. Following [18], we then fit logistic functions (equation 3.12) to the network responses using non-linear least-squares:

$$P(x; \alpha, \beta, \lambda, \gamma) = \gamma + \frac{1 - \gamma - \lambda}{1 + e^{-\beta(x-\alpha)}}. \tag{3.12}$$

Where $P(x)$ is the detection probability (i.e. the probability of the model producing a hit or false alarm), and $\alpha, \beta, \lambda, \gamma$ are free parameters. We used the optimal criterion found in the unperturbed (i.e. control) simulations in the perturbation simulations.

Second, to ensure that our results were not an artefact of the (optimal) criterion we constructed neurometric functions following the procedure described in the supplementary material of [18]. Specifically, for each stimulus intensity, we constructed ROC curves and then computed the AUC (area under the ROC curve), thereby summing over all criteria. To convert the AUC into a quantity comparable to a psychometric function (i.e. so that each neurometric function varied between 0 and 1), we normalised the AUC values, $P_{response} = \mathrm{AUC} - \mathrm{intercept} \times \mathrm{max}$. The intercept was given by the minimium AUC across all conditions, and the max was given by the maximum AUC across all conditions.

3.5 Chapter 3 supplementary material

3.5.1 Apical compartment phase plane

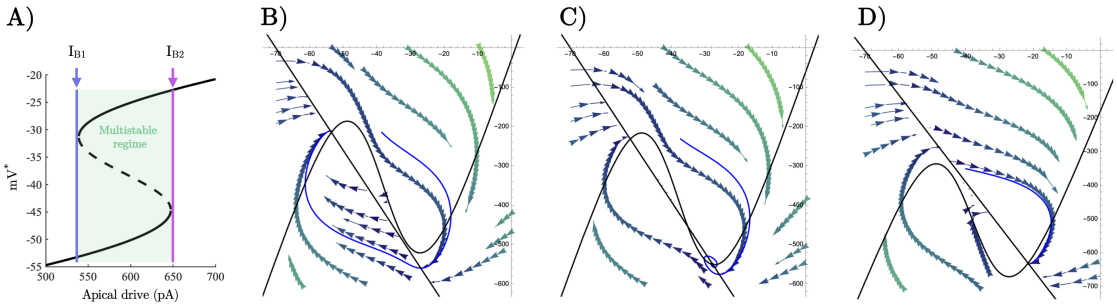

Figure 3.6: **A)** Bifurcation diagram of the $L5_{ET}$ apical compartment. The saddle node bifurcation at I_{B1} generates a stable plateau potential which coexists with the resting state of the apical compartment until the model passes through a second saddle node bifurcation at I_{B2} at which point the resting state of the compartment vanishes and the plateau potential becomes globally attracting. **B-D)** Phase plane representation of the apical compartment showing the nullclines (black) for the following values of the bifurcation parameter; $I_{ext} < I_{B1}$, $I_{B1} < I_{ext} < I_{B2}$, and $I_{ext} > I_{B2}$.

3.5.2 Sweeping the magnitude of model perturbations

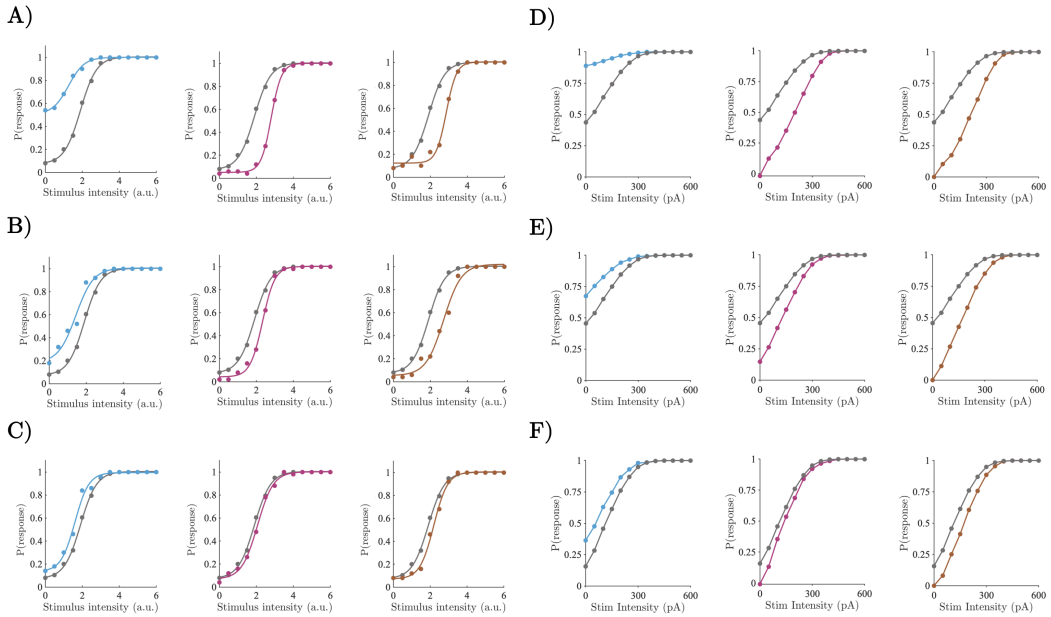

Figure 3.7: **A-C)** Psychometric function fit to spiking model output across apical compartment excitation (blue), apical compartment inhibition (pink), and thalamic inhibition (orange), of varying magnitudes; **A** = 300 pA, **B** = 200 pA, **C** = 100 pA. **D-F)** Same as A-C but for neurometric functions; **D** = 300 pA, **E** = 200 pA, **F** = 100 pA.

3.5.3 Robustness of rivalry duration across burstiness parameters

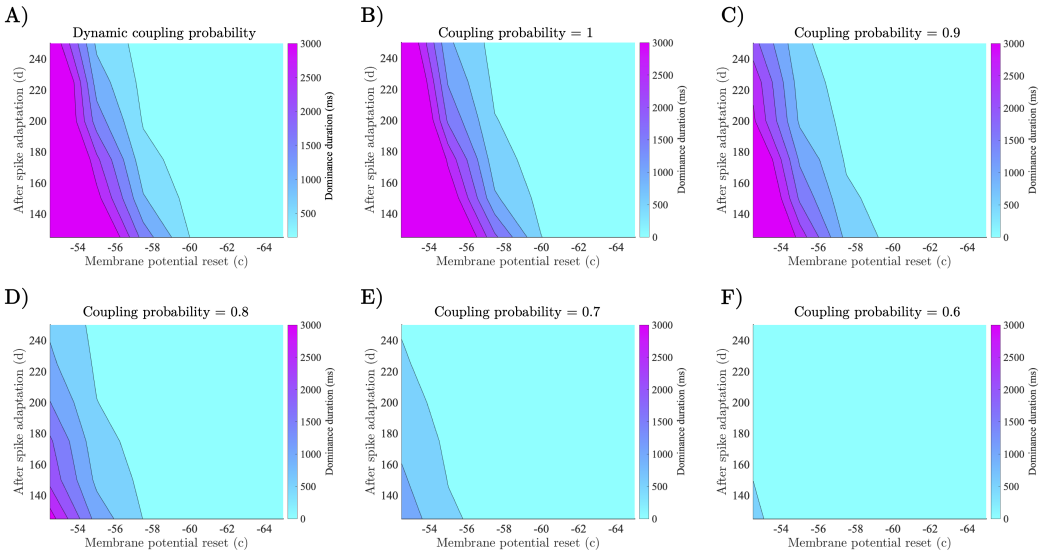

Figure 3.8: **A)** Dominance duration as a function of spike reset values with dynamic inter-compartment coupling (typical values are ~ 0.98 for the dominant population, and ~ 0.1 for the suppressed population; see Figure 3.3D). **B-F)** Dominance duration as a function of spike reset values with stationary coupling probability ranging from 1 to 0.6.

3.5.4 Dynamical regime underlying visual rivalry

To interrogate the structure of the dynamical system underlying the stochastic oscillations, we made inter-compartment coupling deterministic and drove the model with a constant current and asymmetric initial conditions so that the system converged to a state where one of the populations was dominant whilst the other was suppressed. We reasoned that if the oscillations were driven by stochastic jumps between stable fixed points with basins of attraction modulated by adaptation, then in the absence of noise the oscillations should disappear. In contrast, if adaptation exerts a large effect, the oscillations should consist of a stable limit cycle and the model should continue to oscillate in the absence of noise (c.f. [34, 88]). In agreement with the stable limit cycle hypothesis, in the absence of noise the model continued to oscillate (Figure 3.9A). To test the stability of the limit cycle we: i) confirmed the existence of an unstable structure inside the limit cycle; and ii) confirmed that perturbations to the limit cycle decayed back to a stable orbit. With symmetric initial conditions the model converged to a state where excitatory activity on each side of the ring was perfectly matched. Perturbations to this state consisting of a 1 ms pulse of constant drive (50 pA) to the somatic compartment of a single $L5_{ET}$ neuron caused the orbit to converge to the surrounding limit cycle (Figure 3.9B). Such perturbations had little effect on already oscillating orbits, confirming the stability of the limit cycle (Figure 3.9C).

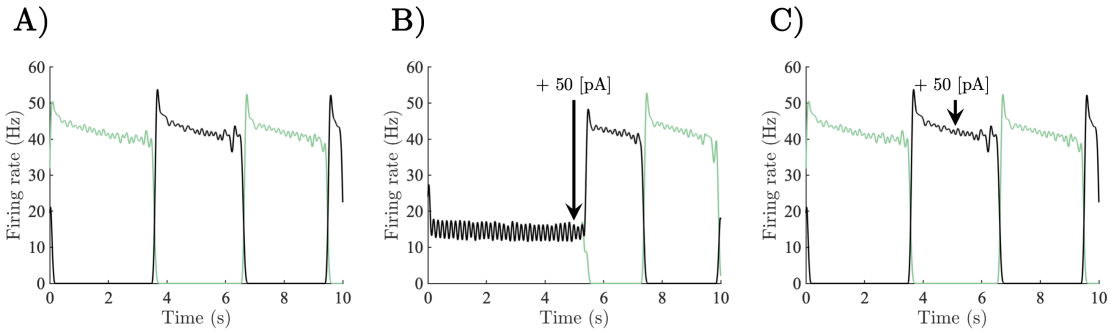

Figure 3.9: **A)** Average firing rate of neuronal populations centred on opposite ends of the ring driven by a constant drive with asymmetric initial conditions. **B)** Average firing rate of neuronal populations simulated with constant drive and symmetric initial conditions. A perturbation was delivered at t = 5000 ms, sending the population orbit to the surrounding stable limit cycle. **C)** Average firing rate of neuronal populations driven by constant drive with asymmetric initial conditions. Perturbation delivered at t = 5000 ms had no substantial effect on the already oscillating orbit, indicative of a stable limit cycle.

3.5.5 Key effects of visual rivalry simulations are preserved in scaled-up model

To help guard against possible biases in the results caused by finite size effects or other simplifying assumptions made in the model, such as the 50/50 excitatory/inhibitory neuron ratio, or the all-to-all connectivity of the cortical ring we constructed a scaled-up version of the model consisting of 2160 neurons (1600 excitatory, 400 inhibitory, 160 thalamic). The scaled-up model had sparse connectivity (12.5% connection probability), and an 80/20 excitatory/inhibitory neuron ratio (i.e., in line with Dale's law). Because of the non-linearities in the model, and the reduction in the number of inhibitory neurons, we could not simply rescale the parameters of the original smaller network. Instead, we retuned the connectivity and adaptation parameters using the procedure described in the methods section above. Parameter values of the scaled-up network model are supplied below in Table 3.3. In line with the behaviour of the small network model reported in the main text, the large network model (Figure 3.10A) generated a Gamma distribution of dominance durations (Figure 3.10B), and was consistent with Levelt's second (Figure 3.10C) and fourth (Figure 3.10D) propositions supporting the robustness of the burst-dependent mechanism of perceptual dominance put forward in the paper. All simulations run in the scaled-up network lasted for 20 s and results were averaged over 20 random seeds.

Parameter	Description	Value	Units
$\lambda_{AMPA}^{E\to E}$	Amplitude of excitatory to excitatory coupling for (AMPA)	$\frac{4.2}{\sqrt{\sigma^{E\to E}2\pi}}$	a.u.
$\lambda_{NMDA}^{E\to E}$	Amplitude of excitatory to excitatory coupling (NMDA)	$\frac{0.84}{\sqrt{\sigma^{E\to E}2\pi}}$	a.u.
$\lambda^{E\to I}$	Amplitude of excitatory to inhibitory coupling (NMDA and AMPA)	$\frac{1.5}{\sqrt{\sigma^{E\to I}2\pi}}$	a.u.
$\lambda^{I\to E}$	Amplitude of inhibitory to excitatory coupling ($GABA_A$)	$\frac{5.85}{\sqrt{\sigma^{I\to E}2\pi}}$	a.u.
$\lambda^{E\to TH}$	Constant excitatory to thalamic coupling constant (AMPA only)	4	a.u.
$\lambda_{AMPA}^{TH\to D}$	Constant thalamic to apical dendrite coupling (AMPA)	10	a.u.
$\lambda_{NMDA}^{TH\to D}$	Constant thalamic to apical dendrite coupling (NMDA)	10	a.u.

Continued on next page

Table 3.2 – continued from previous page

Parameter	Description	Value	Units
$\sigma_{E\to E}$	Spread of excitatory to excitatory coupling	0.25	a.u.
$\sigma_{E\to I}$	Spread of excitatory to inhibitory coupling	2	a.u.
$\sigma_{I\to E}$	Spread of inhibitory to excitatory coupling	2	a.u.
$\tau_{decay:AMPA}$	Decay time of AMPA conductance	6	ms
$\tau_{decay:GABA_A}$	Decay time of $GABA_A$ conductance	6	ms
$\tau_{decay:NMDA}$	Decay time of NMDA conductance	100	ms
$\tau_{coupling}$	Decay time of apical coupling zone	800	ms
τ_{adapt}	Decay time of adaptation current	2000	ms
Δg	Contribution of each spike to adaptation current	0.05	nS
σ_{VR}	Spatial spread of apical drive	20	a.u.
$E_{Excitatory}$	Reverse potential of excitatory synapses	0	mV
$E_{Inhibitory}$	Reverse potential of inhibitory synapses	-75	mV
E_{Adapt}	Reverse potential of adaptation currents.	-80	mV

Table 3.3: Parameter description, values, and units of the (scaled-up) model components described by equations 3.1-3.9.

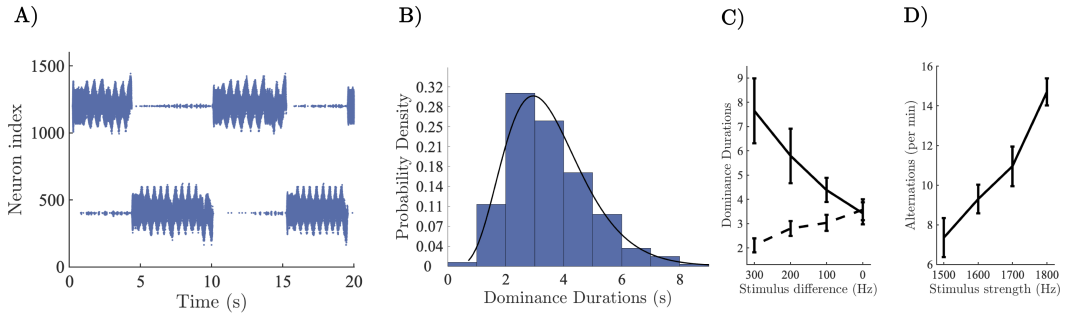

Figure 3.10: **A)** Raster plots of somatic spikes from the scaled-up population of $L5_{ET}$ cells. **B)** Histogram of dominance durations, black line shows the fit of a Gamma distribution with parameters estimated via MLE ($\alpha = 6.2, \theta = 0.56$). **C)** Simulation confirming Levelt's second proposition in scaled-up model. Dashed line shows the dominance duration of the population receiving the decreasing external drive, solid line shows dominance duration of population receiving a fixed drive. **D)** Simulation of Levelt's fourth proposition in scaled-up model.

3.5.6 *In silico* electrophysiology supplemental figures

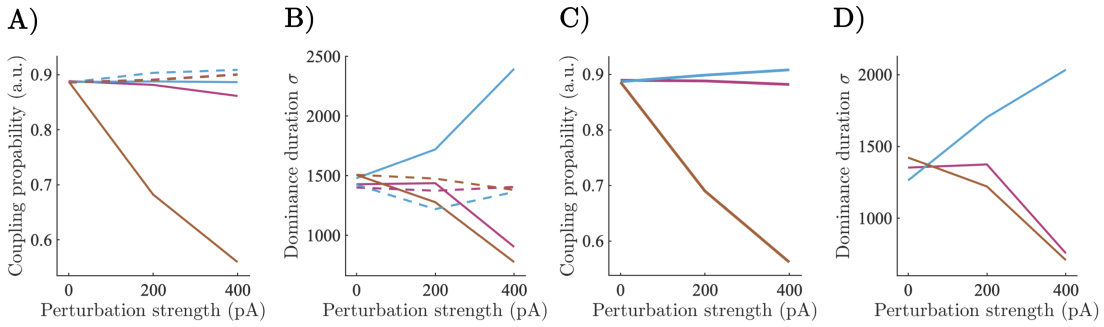

Figure 3.11: **A)** Inter-compartment coupling probability under asymmetric perturbation for perturbed (solid) and unperturbed (dashed) populations as a function of perturbation strength for each perturbation type (colours same as main text). **B)** Dominance duration standard deviation (σ) under asymmetric perturbation as a function of the strength of each perturbation type. **C)** Inter-compartment coupling probability under symmetric perturbations. **D)** Dominance duration standard deviation (σ) under symmetric perturbation as a function of perturbation strength for each perturbation type.

Chapter 3 references

[1] Naoya Takahashi et al. "Active cortical dendrites modulate perception". In: *Science* 354.6319 (2016), pp. 1587–1590.

[2] Naoya Takahashi et al. "Active dendritic currents gate descending cortical outputs in perception". In: *Nature neuroscience* 23.10 (2020), pp. 1277–1285.

[3] C. Qian et al. *Hierarchical and fine-scale mechanisms of binocular rivalry for conscious perception*. bioRxiv. 2023. DOI: 10.1101/2023.02.11.528110.

[4] J. Seo et al. "The thalamocortical inhibitory network controls human conscious perception". In: *Neuroimage* 264 (2022), p. 119748.

[5] Melanie Wilke, K-M Mueller, and David A Leopold. "Neural activity in the visual thalamus reflects perceptual suppression". In: *Proceedings of the National Academy of Sciences* 106.23 (2009), pp. 9465–9470.

[6] Xiao-Jing Wang. *Theoretical Neuroscience: Understanding Cognition*. CRC Press, 2025.

[7] M. Overgaard. *Behavioural Methods in Consciousness Research*. Oxford University Press, 2015.

[8] Michael A. Pitts, Scott Metzler, and Steven A. Hillyard. "Isolating neural correlates of conscious perception from neural correlates of reporting one's perception". In: *Frontiers in Psychology* 5 (2014). DOI: 10.3389/fpsyg.2014.01078.

[9] Claire Sergent, Sylvain Baillet, and Stanislas Dehaene. "Timing of the brain events underlying access to consciousness during the attentional blink". In: *Nature Neuroscience* 8.10 (2005), pp. 1391–1400. DOI: 10.1038/nn1549.

[10] Selene Ciceri et al. "The neural and computational architecture of feedback dynamics in mouse cortex during stimulus report". In: *bioRxiv* (2024). DOI: 10.1101/2023.07.19.549692.

[11] S.D. Gale et al. "Backward masking in mice requires visual cortex". In: *Nature Neuroscience* 27 (2024), pp. 129–136.

[12] M. N. Oude Lohuis et al. "Multisensory task demands temporally extend the causal requirement for visual cortex in perception". In: *Nature Communications* 13.1 (2022), p. 2864. DOI: 10.1038/s41467-022-30600-4.

[13] G. Palagina, J.F. Meyer, and S.M. Smirnakis. "Complex Visual Motion Representation in Mouse Area V1". In: *The Journal of Neuroscience* 37 (2017), pp. 164–183.

[14] B.J. He. "Towards a pluralistic neurobiological understanding of consciousness". In: *Trends in Cognitive Sciences* 27 (2023), pp. 420–432.

[15] J. Aru et al. "Coupling the State and Contents of Consciousness". In: *Frontiers in Systems Neuroscience* 13 (2019), p. 43.

[16] J. Aru, M. Suzuki, and M.E. Larkum. "Cellular Mechanisms of Conscious Processing". In: *Trends in Cognitive Sciences* 24 (2020), pp. 814–825.

[17] T. Bachmann, M. Suzuki, and J. Aru. "Dendritic integration theory: A thalamo-cortical theory of state and content of consciousness". In: *Philosophical Inquiries* 1 (2020), 10.33735/phimisci.2020.ii.52.

[18] N. Takahashi et al. "Active Cortical Dendrites Modulate Perception". In: *Science* 354 (2016), pp. 1587–1590.

[19] N. Takahashi et al. "Active Dendritic Currents Gate Descending Cortical Outputs in Perception". In: *Nature Neuroscience* (2020). DOI: 10.1038/s41593-020-0677-8.

[20] C. J. Whyte et al. "Thalamic contributions to the state and contents of consciousness". In: *Neuron* 112.10 (2024), pp. 1611–1625. DOI: 10.1016/j.neuron.2024.04.019.

[21] Tomá Marvan et al. “Apical amplification-A cellular mechanism of conscious perception?” In: *Neuroscience of Consciousness* 2021.2 (2021), niab036. DOI: 10.1093/nc/niab036.

[22] J.M. Phillips, N.A. Kambi, and Y.B. Saalmann. “A Subcortical Pathway for Rapid, Goal-Driven, Attentional Filtering”. In: *Trends in Neurosciences* 39 (2016), pp. 49–51.

[23] David Alais and Randolph Blake. *Binocular Rivalry*. MIT Press, 2005.

[24] Randolph Blake and Frank Tong. “Binocular rivalry”. In: *Scholarpedia* 3.12 (2008), p. 1578. DOI: 10.4249/scholarpedia.1578.

[25] David Carmel et al. “How to Create and Use Binocular Rivalry”. In: *Journal of Visualized Experiments* 45 (2010), p. 2030. DOI: 10.3791/2030.

[26] Frank Tong, Ming Meng, and Randolph Blake. “Neural bases of binocular rivalry”. In: *Trends in Cognitive Sciences* 10.11 (2006), pp. 502–511. DOI: 10.1016/j.tics.2006.09.003.

[27] Jean-Michel Hupé, C. M. Signorelli, and D. Alais. “Two paradigms of bistable plaid motion reveal independent mutual inhibition processes”. In: *Journal of Vision* 19.4 (2019), p. 5. DOI: 10.1167/19.4.5.

[28] Daria Bogatova, Stelios M. Smirnakis, and Galina Palagina. “Tug-of-Peace: Visual Rivalry and Atypical Visual Motion Processing in MECP2 Duplication Syndrome of Autism”. In: *Eneuro* 11.1 (2024), ENEURO.0102–23.2023. DOI: 10.1523/ENEURO.0102-23.2023.

[29] Stephen Grossberg et al. “How does binocular rivalry emerge from cortical mechanisms of 3-D vision?” In: *Vision Research* 48.21 (2008), pp. 2232–2250. DOI: 10.1016/j.visres.2008.06.024.

[30] Carlo R. Laing, Thomas Frewen, and Ioannis G. Kevrekidis. “Reduced models for binocular rivalry”. In: *Journal of Computational Neuroscience* 28.3 (2010), pp. 459–476. DOI: 10.1007/s10827-010-0227-6.

[31] Carlo R. Laing and Carson C. Chow. “A Spiking Neuron Model for Binocular Rivalry”. In: *Journal of Computational Neuroscience* 12.1 (2002), pp. 39–53. DOI: 10.1023/A:1014942129705.

[32] S. Safavi and P. Dayan. “Multistability, perceptual value, and internal foraging”. In: *Neuron* 110.18 (2022), 3020–3034.e6. DOI: 10.1016/j.neuron.2022.07.024.

[33] A. Shpiro et al. “Dynamical Characteristics Common to Neuronal Competition Models”. In: *Journal of Neurophysiology* 97.1 (2007), pp. 462–473. DOI: 10.1152/jn.00604.2006.

[34] A. Shpiro et al. “Balance between noise and adaptation in competition models of perceptual bistability”. In: *Journal of Computational Neuroscience* 27.1 (2009), pp. 37–54. DOI: 10.1007/s10827-008-0125-3.

[35] Hugh R. Wilson. “Computational evidence for a rivalry hierarchy in vision”. In: *Proceedings of the National Academy of Sciences* 100.24 (2003), pp. 14499–14503. DOI: 10.1073/pnas.2333622100.

[36] Hugh R. Wilson. “Minimal physiological conditions for binocular rivalry and rivalry memory”. In: *Vision Research* 47.21 (2007), pp. 2741–2750. DOI: 10.1016/j.visres.2007.07.007.

[37] Hugh R. Wilson. “Binocular contrast, stereopsis, and rivalry: Toward a dynamical synthesis”. In: *Vision Research* 140 (2017), pp. 89–95. DOI: 10.1016/j.visres.2017.07.016.

[38] Eugene M. Izhikevich. “Simple model of spiking neurons”. In: *IEEE Transactions on Neural Networks* 14.6 (2003), pp. 1569–1572. DOI: 10.1109/TNN.2003.820440.

[39] Eugene M. Izhikevich. "Which model to use for cortical spiking neurons?" In: *IEEE Transactions on Neural Networks* 15.5 (2004), pp. 1063–1070. DOI: 10.1109/TNN.2004.832719.

[40] Eugene M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, 2006. DOI: 10.7551/mitpress/2526.001.0001.

[41] Richard Naud and Henning Sprekeler. "Sparse bursts optimize information transmission in a multiplexed neural code". In: *Proceedings of the National Academy of Sciences* 115.27 (2018), E6329–E6338. DOI: 10.1073/pnas.1720995115.

[42] M.E. Larkum. "Top-down Dendritic Input Increases the Gain of Layer 5 Pyramidal Neurons". In: *Cerebral Cortex* 14 (2004), pp. 1059–1070.

[43] Matthew E. Larkum. "Are Dendrites Conceptually Useful?" In: *Neuroscience* 489 (2022), pp. 4–14. DOI: 10.1016/j.neuroscience.2022.03.008.

[44] M. E. Larkum et al. "Synaptic Integration in Tuft Dendrites of Layer 5 Pyramidal Neurons: A New Unifying Principle". In: *Science* 325.5941 (2009), pp. 756–760. DOI: 10.1126/science.1171958.

[45] Brandon R. Munn et al. "A thalamocortical substrate for integrated information via critical synchronous bursting". In: *Proceedings of the National Academy of Sciences* 120.46 (2023), e2308670120. DOI: 10.1073/pnas.2308670120.

[46] Brandon R. Munn et al. "Neuronal connected burst cascades bridge macroscale adaptive signatures across arousal states". In: *Nature Communications* 14.1 (2023), Article 1. DOI: 10.1038/s41467-023-42465-2.

[47] Richard Naud, Brice Bathellier, and Wulfram Gerstner. "Spike-timing prediction in cortical neurons with active dendrites". In: *Frontiers in Computational Neuroscience* 8 (2014). DOI: 10.3389/fncom.2014.00090.

[48] M. Suzuki and M.E. Larkum. "General Anesthesia Decouples Cortical Pyramidal Neurons". In: *Cell* 180 (2020), 666–676.e13.

[49] A. S. Shai et al. "Physiology of Layer 5 Pyramidal Neurons in Mouse Primary Visual Cortex: Coincidence Detection through Bursting". In: *PLOS Computational Biology* 11.3 (2015), e1004090. DOI: 10.1371/journal.pcbi.1004090.

[50] Z. Wang, W. Dai, and D. W. McLaughlin. "Ring models of binocular rivalry and fusion". In: *Journal of Computational Neuroscience* 48.2 (2020), pp. 193–211. DOI: 10.1007/s10827-020-00744-7.

[51] Y. Zerlaut et al. "Modeling mesoscopic cortical dynamics using a mean-field model of conductance-based networks of adaptive exponential integrate-and-fire neurons". In: *Journal of Computational Neuroscience* 44.1 (2018), pp. 45–61. DOI: 10.1007/s10827-017-0668-2.

[52] R. A. Mease and A. J. Gonzalez. "Corticothalamic Pathways From Layer 5: Emerging Roles in Computation and Pathology". In: *Frontiers in Neural Circuits* 15 (2021), p. 730211. DOI: 10.3389/fncir.2021.730211.

[53] G.M.G. Shepherd and N. Yamawaki. "Untangling the Cortico-Thalamo-Cortical Loop: Cellular Pieces of a Knotty Circuit Puzzle". In: *Nature Reviews Neuroscience* (2021). DOI: 10.1038/s41583-021-00459-3.

[54] Alain Destexhe. "Self-sustained asynchronous irregular states and Up-Down states in thalamic, cortical and thalamocortical networks of nonlinear integrate-and-fire neurons". In: *Journal of Computational Neuroscience* 27.3 (2009), pp. 493–506. DOI: 10.1007/s10827-009-0164-4.

[55] Kenneth H. Britten et al. "The analysis of visual motion: A comparison of neuronal and psychophysical performance". In: *The Journal of Neuroscience* 12.12 (1992), pp. 4745–4765. DOI: 10.1523/JNEUROSCI.12-12-04745.1992.

[56] Nikos K Logothetis, David A Leopold, and DL Sheinberg. "What is rivalling during binocular rivalry?" In: *Nature* 380.6575 (1996), pp. 621–624.

[57] D.A. McCormick. "Cholinergic and Noradrenergic Modulation of Thalamocortical Processing". In: *Trends in Neurosciences* 12 (1989), pp. 215–221.

[58] H. R. Wilson and J. D. Cowan. "Evolution of the Wilson-Cowan equations". In: *Biological Cybernetics* 115.6 (2021), pp. 643–653. DOI: 10.1007/s00422-021-00912-7.

[59] Hsin-Hung Li et al. "Attention model of binocular rivalry". In: *Proceedings of the National Academy of Sciences* 114.30 (2017), E6192–E6201.

[60] Brian E Kalmbach et al. "Signature morpho-electric, transcriptomic, and dendritic properties of human layer 5 neocortical pyramidal neurons". In: *Neuron* 109.18 (2021), pp. 2914–2927.

[61] Jan W. Brascamp, P. Christiaan Klink, and Willem J. M. Levelt. "The 'laws' of binocular rivalry: 50 years of Levelt's propositions". In: *Vision Research* 109 (2015), pp. 20–37. DOI: 10.1016/j.visres.2015.02.019.

[62] W. J. M. Levelt. "Note on the Distribution of Dominance Times in Binocular Rivalry". In: *British Journal of Psychology* 58.1-2 (1967), pp. 143–145. DOI: 10.1111/j.2044-8295.1967.tb01068.x.

[63] Steven Strogatz. *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Second. CRC Press Taylor & Francis Group, 2018.

[64] N. K. Logothetis and D. A. Leopold. "What is rivalling during binocular rivalry?" In: *Nature* 380 (1996), pp. 621–624.

[65] Yoram S. Bonneh et al. "Motion-Induced Blindness and Troxler Fading: Common and Different Mechanisms". In: *PLoS ONE* 9.3 (2014), e92894. DOI: 10.1371/journal.pone.0092894.

[66] Jan W. Brascamp et al. "The time course of binocular rivalry reveals a fundamental role of noise". In: *Journal of Vision* 6.11 (2006), pp. 8–8. DOI: 10.1167/6.11.8.

[67] A. Buckthought, J. Kim, and H. R. Wilson. "Hysteresis effects in stereopsis and binocular rivalry". In: *Vision Research* 48.6 (2008), pp. 819–830. DOI: 10.1016/j.visres.2007.12.013.

[68] Ming Meng and Frank Tong. "Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures". In: *Journal of Vision* 4.7 (2004), p. 2. DOI: 10.1167/4.7.2.

[69] J. Poort and A. F. Meyer. "Vision: Depth perception in climbing mice". In: *Current Biology* 31.10 (2021), R486–R488. DOI: 10.1016/j.cub.2021.03.066.

[70] W.J.M. Levelt. "On binocular rivalry". PhD thesis. Soesterberg, The Netherlands: Institute for Perception RVO-TNO, 1965.

[71] M. Wilke, K.-M. Mueller, and D.A. Leopold. "Neural activity in the visual thalamus reflects perceptual suppression". In: *Proceedings of the National Academy of Sciences of the United States of America* 106 (2009), pp. 9465–9470.

[72] Eric M Trautmann et al. "Large-scale high-density brain-wide neural recording in nonhuman primates". In: *Nature Neuroscience* (2025), pp. 1–14.

[73] E.J. Müller et al. "Core and Matrix Thalamic Sub-Populations Relate to Spatio-Temporal Cortical Connectivity Gradients". In: *Neuroimage* 222 (2020), p. 117224.

[74] E.J. Müller et al. "The Non-Specific Matrix Thalamus Facilitates the Cortical Information Processing Modes Relevant for Conscious Awareness". In: *Cell Reports* 42 (2023), p. 112844.

[75] Jorge Jaramillo, Jorge F Mejias, and Xiao-Jing Wang. "Engagement of pulvino-cortical feedforward and feedback pathways in cognitive computations". In: *Neuron* 101.2 (2019), pp. 321–336.

[76] C. Fernandez Pujol, E.G. Blundon, and A.R. Dykstra. "Laminar specificity of the auditory perceptual awareness negativity: A biophysical modeling study". In: *PLoS Computational Biology* 19 (2023), e1011003.

[77] R.J. Douglas and K.A.C. Martin. "Neuronal circuits of the neocortex". In: *Annual Review of Neuroscience* 27 (2004), pp. 419–451.

[78] K.D. Harris and G.M.G. Shepherd. "The neocortical circuit: themes and variations". In: *Nature Neuroscience* 18 (2015), pp. 170–181.

[79] Farzin Tahvili and Alain Destexhe. "A mean-field model of gamma-frequency oscillations in networks of excitatory and inhibitory neurons". In: *Journal of Computational Neuroscience* 52.2 (2024), pp. 165–181.

[80] Abhilash Dwarakanath et al. "Prefrontal state fluctuations control access to consciousness". In: *bioRxiv* (2020).

[81] V. Kapoor et al. "Decoding internally generated transitions of conscious contents in the prefrontal cortex without subjective reports". In: *Nature Communications* 13 (2022), p. 1535.

[82] Eugene M. Izhikevich and Gerald M. Edelman. "Large-scale model of mammalian thalamocortical systems". In: *Proceedings of the National Academy of Sciences* 105.9 (2008), pp. 3593–3598. DOI: 10.1073/pnas.0712231105.

[83] Renaud Greget et al. "Simulation of postsynaptic glutamate receptors reveals critical features of glutamatergic transmission". In: *PLoS One* 6.12 (2011), e28380.

[84] Alexander Naka and Hillel Adesnik. "Inhibitory circuits in cortical layer 5". In: *Frontiers in neural circuits* 10 (2016), p. 35.

[85] Peter Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2005.

[86] Wulfram Gerstner et al. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. 1st. Cambridge University Press, 2014. DOI: 10.1017/CBO9781107447615.

[87] Eugene M Izhikevich. "Hybrid spiking models". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368.1930 (2010), pp. 5061–5070.

[88] Rubén Moreno-Bote, John Rinzel, and Nava Rubin. "Noise-induced alternations in an attractor network model of perceptual bistability". In: *Journal of neurophysiology* 98.3 (2007), pp. 1125–1139.

Part 2

Dynamical Principles of Awareness and Suppression

Chapter 4

From Cellular Neurobiology to Psychophysics

> All models are wrong, because all models ignore some details. All models are also wrong because they represent only a biased view of the processes they claim to capture. And all models are wrong because they favour simplicity over absolute accuracy. All models are wrong the same way poems are wrong; they capture an essence, if not a perfect literal truth.
>
> — *Grace Lindsay*

If the models in part 1 of this thesis are wrong [1], then the models I explore in part 2 are even more wrong. These new models average over, or idealise away, much of the hard-won neurobiological complexity we explored in the previous chapters. Although, despite being very wrong indeed, the models are arguably a beautiful piece of poetry. They capture the essence of a phenomenon, in this case, visual awareness and suppression, in a way that the more complex model we explored in the previous chapter will never be able to. They are also simple enough that it is possible to derive closed-form expressions, reminiscent of what we see in the physical sciences, that even my admittedly simple mind can grasp. The existence of these expressions provides a tantalising hint that there might be a core set of principles governing perceptual awareness that are out there to be discovered, and perhaps even understood.

In saying that the models we study in this section average over or idealise away neurobiological complexity, I am by no means implying that these details are not important. They are important, which is why I spent the majority of my PhD painstakingly building them into a spiking model. My life would have been a lot easier, albeit a lot less interesting, if they could have been ignored. This is not to say, however, that such details are always essential. Indeed, for the purposes of chapters 5 and 6, they are not essential and are therefore left out. For the sake of conceptual continuity, here I show how one can derive the simple, purely cortical, neural mass model of visual rivalry studied in part two of this thesis from a more complex neural mass model that more faithfully represents the details of the full thalamocortical dynamical system explored in previous chapters. Through numerical simulation, I show that the purely cortical neural mass model of visual rivalry has dynamics that are nearly identical to the

more detailed thalamocortical model. Further, I show that the closed-form expression for dominance duration derived from the reduced model still accurately describes the behaviour of the full thalamocortical system, validating the assumptions underlying its reduction.

4.1 From thalamocortical dynamics to cortical masses: the case of binocular rivalry

Minimally, to reproduce the nonlinear oscillation between mutually exclusive percepts characteristic of binocular rivalry, a model must have the following three elements. First, separate stimulus-selective excitatory cortical populations must be driven by independent monocular inputs from the left and right eyes. Second, the populations must be coupled through strong competitive inhibition (resulting in mutually exclusive percepts). And third, the excitatory cortical populations must slowly adapt over time, eventually releasing the suppressed population from inhibition, resulting in a perceptual switch [2, 3].

To derive a thalamocortical neural mass model with these dynamical properties, I assume that neural populations within cortex and thalamus can be divided into reciprocally-coupled populations of excitatory and inhibitory cells, and that the populations are accurately described by the Wilson-Cowan equations [4, 5]. To further simplify the problem, I make three additional assumptions. One, competitive inhibition between monocular populations is primarily cortical. Two, the rapid dynamics of inhibitory cells are well approximated by their equilibrium values. And three, the influence of higher order thalamus on cortex is primarily mediated by input to metabotropic receptors on the apical coupling zone of $L5_{ET}$ cells (i.e., on the oblique dendrites) which effectively controls the gain of the cells firing rate (see figure 3.1D and [6]). These assumptions in hand, we arrive at the following eight-dimensional system of ordinary differential equations (ODEs) describing the dynamics of two competing matrix thalamus - $L5_{ET}$ thalamocortical neural masses:

$$
\begin{aligned}
\tau_E \dot{E}_L &= -E_L + (\alpha + U_L) f(L + \varepsilon E_L - a E_R - g H_L) \\
\tau_{Ud} \dot{U}_L &= -U_L + \frac{\tau_{Ud}}{\tau_{Ur}} (1 - U_L) T_L \\
\tau_E \dot{T}_L &= -T_L + f(E_L) \\
\tau_H \dot{H}_L &= -H_L + E_L \\
\tau_E \dot{E}_R &= -E_R + (\alpha + U_R) f(R + \varepsilon E_R - a E_L - g H_R) \\
\tau_{Ud} \dot{U}_R &= -U_R + \frac{\tau_{Ud}}{\tau_{Ur}} (1 - U_R) T_R \\
\tau_E \dot{T}_R &= -T_R + f(E_R) \\
\tau_H \dot{H}_R &= -H_R + E_R.
\end{aligned}
\tag{4.1}
$$

The eight state variables $E_{L,R}$, $U_{L,R}$, $T_{L,R}$, and $H_{L,R}$, represent the aggregate effects of; excitatory activity of orientation selective monocular populations in cortical $L5_B$ (driven by L and R representing monocular input from each eye); postsynaptic dynamics of metabolic receptors (mGluR/mACh; [7]) which control the gain of the $L5_{ET}$ cells' transfer function; excitatory activity of visual higher-order thalamus (i.e. the pulvinar); and the slow hyperpolarising adaptation current of each cortical population. The population transfer function is a simple threshold nonlinearity $f(x) = \max(x, 0)$, which for $L5_{ET}$ cells, provides a good description of the single cell F/I curves (again, see figure 3.1D and [6]). The parameters α, ε, a, and g,

are dimensionless and represent the baseline gain of the cortical transfer function, and the strength of recurrent excitation, competitive inhibition, and adaptation, respectively. The time constants τ_E, τ_{Ud}, τ_{Ur}, and τ_H are in units of milliseconds. Based on the relatively rapid rise time of metabolic glutamate receptors [8], we use $\tau_{Ur} = 70$ ms as a conservative estimate, receptor dynamics can be well approximated by their equilibrium values which can then be treated as a parameter $\beta = \alpha + U_{L,R}(\infty)$. This allows us to reduce our system of eight ODEs to the following four-dimensional purely cortical system studied by Wilson [3]:

$$\begin{aligned}
\tau_E \dot{E}_L &= -E_M + \beta f(L + \varepsilon E_L - a E_R - g H_L) \\
\tau_H \dot{H}_L &= -H_L + E_L \\
\tau_E \dot{E}_R &= -E_R + \beta f(R + \varepsilon E_R - a E_L - g H_L) \\
\tau_H \dot{H}_R &= -H_R + E_R.
\end{aligned} \tag{4.2}$$

Importantly, the simplicity of this system allows for the derivation of a closed-form expression for the dominance duration of the percept represented by each monocular population as a function of model parameters (see Wilson [3] for details). We show the expression for the population driven by input from the left eye below:

$$T_L = \tau_H \ln\left(\frac{1}{\beta + g - \varepsilon - a(L/R)}\right). \tag{4.3}$$

The comparable expression for the right monocular population can be obtained by flipping the ratio of stimulus strengths $L/R \to R/L$. Equation 4.3 captures Levelts second law (described in chapter 3), via the dependence on the ratio of stimulus strengths in the denominator [3]. In addition, with minor modification of the population transfer function [9], the expression can be extended to account for Levelts fourth law (again described in chapter 3). Numerical simulations show that the dynamics of each model are near identical (figure 4.1A-B). In addition, the closed-form expression for dominance duration given by equation 4.3 provides an excellent description of both models across a wide range of parameters, validating the assumptions made in the reduction (figure 4.1C). Parameter values and simulation details are given in the supplementary material at the end of the chapter. The validity of this result is significant. As we discussed in the introduction to this thesis, visual rivalry is the only perceptual awareness paradigm that is well enough understood to be characterised by simple psychophysical laws. When building complex models, such constraints are crucial, indeed showing that the multicompartment thalamocortical spiking model, presented in chapter 3, was consistent with these laws provided a first test of the face validity of the model and gave us sufficient confidence for it to be worthwhile to generate additional neuronal predictions.

In the next two chapters, we explore the dynamical principles underlying a recent extension of binocular rivalry, tracking continuous flash suppression (tCFS), which allows experimenters to reliably measure the threshold for visual awareness and suppression. We extend Wilson's (2007) minimal model of binocular rivalry to tCFS with the hope of eventually building a full multicompartment thalamocortical spiking model of the task that we can constrain with psychophysical laws as we did with binocular rivalry. In the meantime, however, we have a lot of psychophysics and dynamical modelling to do to figure out whether such psychophysical laws exist in visual awareness paradigms outside of binocular rivalry. If we can show that such laws exist, the equivalence of the thalamocortical mass and the reduced cortical mass should give us a degree of confidence that it will be possible to build and validate microscale spiking models of this new paradigm that can, in the fullness of time, be tested in animal models

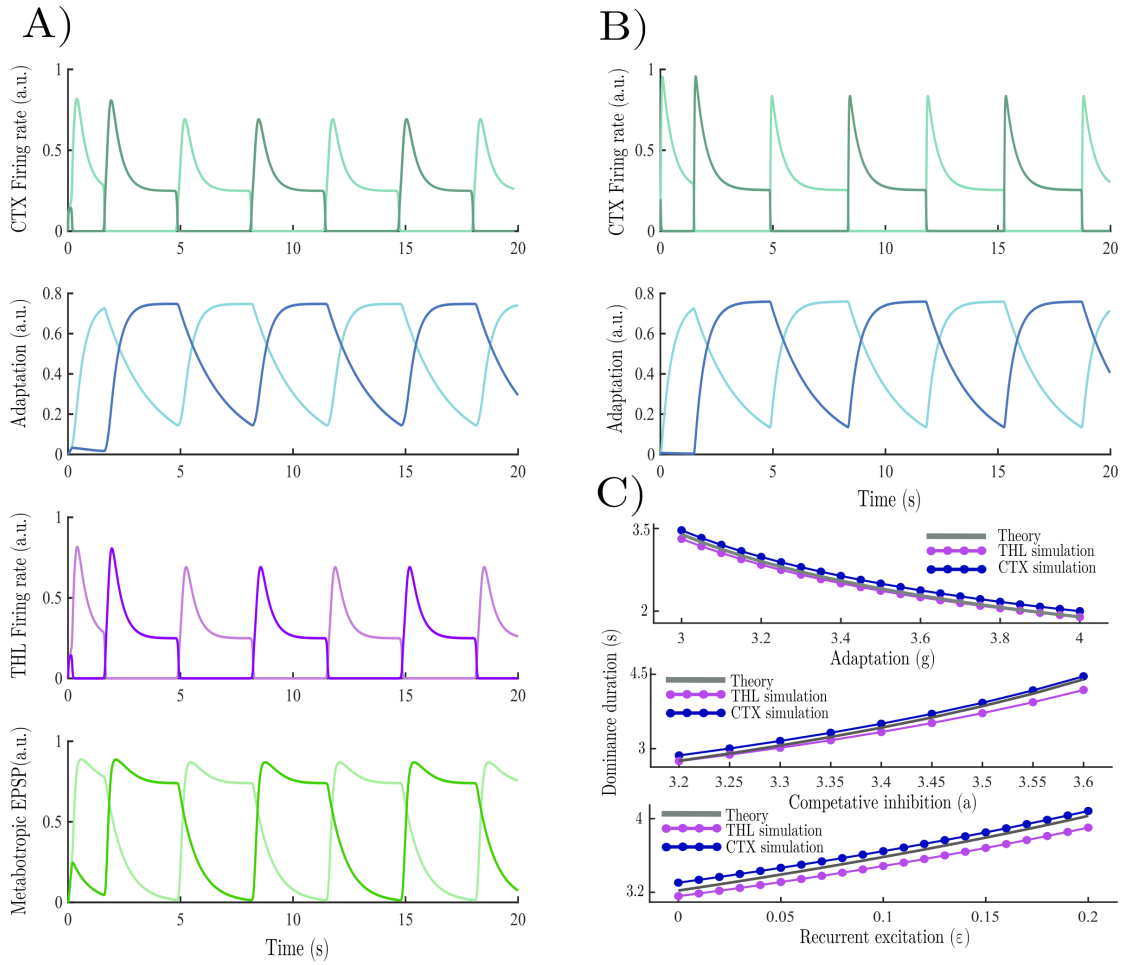

Figure 4.1: **A)** Simulation of the full thalamocortical model for each monocular population which includes cortical firing rates (upper), adaptation current dynamics (middle upper), thalamic firing rates (middle lower) and metabotropic EPSPs (lower). **B)** Simulation of the reduced cortical model only includes firing rates (upper) and adaptation current dynamics (lower). **C)** Comparison of average dominance durations from the full thalamocortical model (purple), purely cortical model (blue) and closed-form theoretical prediction from equation 3 (grey) under parameter sweeps including, adaptation (upper), competitive inhibition (middle) and recurrent excitation (lower).

and potentially provide us with a general cellular-level explanation of the threshold for visual awareness and suppression.

4.2 When do neurobiological details matter for psychology?

Before we turn to the next chapter, I want to address a question which I suspect will be at the forefront of readers' minds. Specifically, when do neurobiological details matter and when can they be idealised away? Here I give a brief and somewhat pragmatic answer that will do for the purposes of the next two chapters. Neurobiological detail matters when they are experimentally manipulated. Simply put, it is not possible to simulate an experimental intervention onto a variable if it is not included in a model. In addition, perturbating a system will, in all likelihood, move variables far away from their equilibrium values, violating the assumptions underlying the reduction of the thalamocortical model to the purely cortical

model. The experiments we simulate in the next two chapters are purely behavioural, which allows me to treat the thalamocortical state variables as parameters. I will return to discuss the role of neurobiological complexity in computational modelling in more detail in the conclusion of the thesis.

4.3 Chapter 4 supplementary material

4.3.1 Model parameters

All simulations were run in MATLAB 2023b using the ode45 function to implement Runge-Kutta numerical integration using the parameter values in table 4.3.1.

Parameter	Description	Value	Units
τ_E	Neuronal population time constant	15	ms
τ_H	Hyperpolarizing adaptation current time constant	1000	ms
τ_{Ur}	Metabotropic receptor dynamics rise time	70	ms
τ_{Ud}	Metabotropic receptor dynamics decay time	800	ms
L	Strength of input from left eye	1	a.u.
R	Strength of input from right eye	1	a.u.
ε	Strength of excitatory recurrent projections	0–0.2 (0.05)	a.u.
a	Strength of competitive inhibition	3.3–3.5 (3.4)	a.u.
g	Strength of adaptation current	3–4 (3)	a.u.
α	Baseline gain of cortical population transfer function	0.2	a.u.
β	Stationary gain of cortical population transfer function	1	a.u.

Table 4.1: Parameter description, values, and units of the two models. Base parameter values used in parameter sweeps shown in parentheses.

Chapter 4 references

[1] Grace Lindsay. *Models of the mind: how physics, engineering and mathematics have shaped our understanding of the brain*. Bloomsbury Publishing, 2021.

[2] A. Shpiro et al. "Dynamical Characteristics Common to Neuronal Competition Models". In: *Journal of Neurophysiology* 97.1 (2007), pp. 462–473. DOI: 10.1152/jn.00604.2006.

[3] Hugh R. Wilson. "Minimal physiological conditions for binocular rivalry and rivalry memory". In: *Vision Research* 47.21 (2007), pp. 2741–2750. DOI: 10.1016/j.visres.2007.07.007.

[4] H. R. Wilson and J. D. Cowan. "Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons". In: *Biophysical Journal* 12.1 (1972), pp. 1–24. DOI: 10.1016/S0006-3495(72)86068-5.

[5] H. R. Wilson and J. D. Cowan. "Evolution of the WilsonCowan equations". In: *Biological Cybernetics* 115.6 (2021), pp. 643–653. DOI: 10.1007/s00422-021-00912-7.

[6] M.E. Larkum. "Top-down Dendritic Input Increases the Gain of Layer 5 Pyramidal Neurons". In: *Cerebral Cortex* 14 (2004), pp. 1059–1070.

[7] J. Aru, M. Suzuki, and M.E. Larkum. "Cellular Mechanisms of Conscious Processing". In: *Trends in Cognitive Sciences* 24 (2020), pp. 814–825.

[8] T. Kukaj et al. "Kinetic fingerprinting of metabotropic glutamate receptors". In: *Communications Biology* 6.1 (2023), p. 104. DOI: 10.1038/s42003-023-04468-z.

[9] J. Seely and C. C. Chow. "Role of mutual inhibition in binocular rivalry". In: *Journal of Neurophysiology* 106.5 (2011), pp. 2136–2150. DOI: 10.1152/jn.00228.2011.

Chapter 5

A Minimal Physiological Model of Perceptual Suppression and Breakthrough in Visual Rivalry

> One of the pleasures of looking at the world through mathematical eyes is that you can see certain patterns that would otherwise be hidden.
>
> — *Steven Strogatz*

The next two chapters extend an established, analytically-tractable model of binocular rivalry to two complementary paradigms: a threshold detection variant of binocular rivalry, and a new variant of continuous flash suppression (tracking continuous flash suppression; tCFS), which provide complementary measures of the temporal evolution, and magnitude, of the contrast threshold required for a stimulus to be suppressed from awareness, and to break through suppression into awareness. Chapters 5 and 6 were written as a pair of companion papers, but were published independently. Both chapters, therefore, cover much of the same ground but in different levels of detail and with different aims and audiences in mind. In chapter 5, we use numerical simulations to qualitatively explore the mesoscale physiological mechanisms underlying perceptual suppression (and breakthrough) in binocular rivalry and tCFS, revealing a shared mechanism of perceptual suppression. This chapter is aimed at experimental psychologists and therefore suppresses much of the mathematical detail. Having established the validity of the model and explored conceptual connections to similar well-established paradigms, in chapter 6, which is aimed at computational and theoretical neuroscientists, we leverage the simplicity of the model to quantitatively characterise the threshold for suppression and breakthrough in tCFS.

5.1 Chapter 5 Introduction

When one eye is presented with an image that is different from the other eye's image, the normal process of binocular fusion is prevented [1–4]. Instead, only one eyes image is seen while the other image is suppressed from awareness by a process known as interocular suppression.

An interesting feature of this paradigm is that the consciously-perceived image alternates over time as the two images compete in a kind of rivalry for perceptual dominance, with perceptual switches occurring every second or two in a stochastic manner.

This process is traditionally studied using binocular rivalry (Fig 5.1A; [3, 5, 6]), a paradigm with a long experimental history dating back over a century [7, 8] with a rich literature in psychophysics [9, 10], neuroimaging [11–16] and non-human primate electrophysiology [17–23]. Binocular rivalry received a boost in popularity in the 1990s, when wide spread interest in the experimental study of consciousness emerged and binocular rivalry was proposed as a key empirical paradigm [21, 24]. The appeal was simple: in rivalry, two visual stimuli enter the visual system, but only one is consciously experienced. This offered a practical way to study several interesting questions central to the neuroscience of consciousness: Can a stimulus suppressed from awareness still guide behaviour? Where along the visual pathway does suppression occur? How strong must suppression be to remove a stimulus from awareness?

The question concerning rivalry suppression strength has received considerable attention, with many studies attempting to quantify suppression depth [25–33]. The results of these studies show a range of 28 dB of suppression, with an average suppression depth of approximately 4-5 dB. Thus, while phenomenal experience in rivalry is that of a stimulus completely disappearing from awareness, measures of suppression depth reveal a relatively modest degree of suppression.

Another form of visual rivalry, known as continuous flash suppression (CFS; [34]), has become very popular in recent years [35, 36]. CFS involves presenting a rapid sequence of high-contrast, random patterns to one eye (the mask) and an image to the other that is usually small, low contrast and static (the target). The mask will reliably suppress the target for much longer periods than is typical in binocular rivalry (for example, 10-20 seconds vs 1-2 seconds) and the dependent variable in CFS is usually the time required for the target to emerge from suppression and reach visual awareness.

CFS is claimed to produce stronger suppression than binocular rivalry, although measuring CFS breakthrough times cannot resolve this question. While longer suppression periods in CFS might be due to stronger suppression, it may instead arise from lack of adaptation to the masker due it being continually updated with new random patterns driving unadapted neurons. Due to the fact that binocular rivalry stimuli are sustained, adaptation is thought to reduce suppression depth within each rivalry phase [37]. To determine CFS suppression strength requires a measure of stimulus contrast thresholds rather than breakthrough times. Separate increment thresholds are needed for breakthrough and suppression so that the visible images threshold can be compared with the suppressed images threshold, as done in binocular rivalry experiments.

A recent CFS variant known as tracking CFS (tCFS; [38]) provides a simple method for measuring CFS suppression depth that enables easy comparisons with estimates from binocular rivalry. It involves a target image that continuously changes in log contrast either rising of falling depending on the observers perceptual report (Fig 5.1B). Breakthrough thresholds are registered when a low contrast target that steadily rises in contrast eventually breaks into awareness. As soon as the observer indicates that the target has become visibile, the contrast of the target reverses direction and steadily reduces in contrast until it becomes suppressed once again. As soon as suppression is reported, the target reverses course, and increases in contrast again until breakthrough (and so on, in a cycle). When the cycle ends (e.g., after 10 reversals), the difference between mean thresholds for breakthrough and suppression quantifies

suppression depth.

Due to the relative simplicity of binocular rivalry, and the law like relationship between the physical properties of a stimulus and the duration of the resulting percepts (i.e. Levelts laws; [6, 39]), formal models of binocular rivalry have been proposed, and iteratively refined and tested, for well over half a century [6, 39–41]. This has led to a rich quantitative modelling literature spanning multiple spatial scales and levels of analysis, from implementation level spiking [42–44] and mean field [45–50] models, which describe rivalry dynamics in terms of biophysical processes, to algorithmic and computational level models which cast perceptual switching as optimising Bayesian [51–55] and/or decision theoretic [56, 57] objectives. Quantitative theory and modelling has, however, lagged behind experimental progress in studying variants of visual rivalry (which we use as a catch all term for paradigms that utilise interocular suppression to render stimuli unconscious; e.g. CFS, and binocular rivalry) that allow researchers to quantify the depth of perceptual suppression (for a notable exception see [58]).

In this paper, we take a first step towards developing a quantitative theory of perceptual suppression in visual rivalry by extending an existing model of binocular rivalry to encompass threshold detection variants of binocular rivalry and tCFS. We begin by introducing the minimal model of binocular rivalry proposed by [49]. We chose to start from this model as it shares key features common to (almost) all neuronal models of rivalry, namely strong competitive inhibition and self-adaptation [47]. The model also has sufficient complexity to account for the psychophysical laws known to govern rivalry (Levelts laws; [6, 40]) whilst still being simple enough that it is possible to derive analytic understanding from closed form approximations to the (nonlinear) dynamics of the full network. We use the model to reproduce and explain the results of two experimental paradigms which provide complementary insight into the dynamics of suppression depth in visual rivalry. The first provides an estimate of how suppression depth evolves over time in binocular rivalry [37]. The second, tCFS [38], provides an absolute measure of the magnitude of perceptual suppression by measuring both suppression and breakthrough thresholds.

5.2 Chapter 5 methods

The dynamics of the Wilson model [49] are governed by two key qualitative features. The first — strong inhibition between competing monocular neuronal populations — generates winner-take-all activity, such that only one monocular population can be active at any one time. The second — strong self-adaptation — guarantees that winner-take-all states are transient, even in the absence of noise, by decreasing the activity of the dominant population sufficiently to release the competing population from suppression via a reduction in competitive inhibition. Together these conditions guarantee a nonlinear oscillation, known as a limit cycle, between two mutually exclusive percepts, separated by brief periods of mixed perception, characteristic of binocular rivalry.

The model is described by a system of four nonlinear ordinary differential equations:

$$\begin{aligned}
\tau_E \dot{E}_L &= -E_L + f(L + \varepsilon E_L - a E_R - g_L H_L) \\
\tau_H \dot{H}_L &= -H_L + E_L \\
\tau_E \dot{E}_R &= -E_R + f(R + \varepsilon E_R - a E_L - g_R H_R) \\
\tau_H \dot{H}_R &= -H_R + E_R,
\end{aligned} \tag{5.1}$$

where the population transfer function is the simple threshold linear function $f(x) = \max(x, 0)$, and the four state variables E_L, E_R and H_L, H_R represent the aggregate neuronal activity of competing monocular populations, and the activity dependent hyperpolarising current of each population, respectively. Each neuronal population is driven by separate monocular input (L, R), weakly excited (with strength ε) by recurrent projections within each population, strongly inhibited (with strength a) by the competing population, and undergoes slow self-inhibition (with strength g) via the accumulation of a slow hyperpolarising adaptation current (H_L, H_R) which represents the action of calcium mediated potassium currents [59, 60]. In addition, we note that the model assumes that inhibition is purely a function of excitatory activity and rapidly reaches its asymptotic value allowing the potential six-dimensional system of ODEs to be reduced to four.

Working from the qualitative features described above, [49] derived a minimal set of sufficient conditions for the presence of a limit cycle in the model. Leveraging the fact that $\tau_H \gg \tau_E$ the dynamics of the model can be partitioned into two phases. In the first instantaneous phase $H_L, H_R \approx 0$, and in the second asymptotic phase, $H_L \approx E_L$ and $H_R \approx E_R$. In both partitions the model can be reduced to a two-dimensional system. In the first instantaneous partition, if $E_L, E_R > 0$, equation 6.1 reduces to the following two-dimensional linear system which we write in matrix form:

$$\tau_E \frac{d}{dt} \begin{pmatrix} E_L \\ E_R \end{pmatrix} = \begin{pmatrix} -1+\varepsilon & -a \\ -a & -1+\varepsilon \end{pmatrix} \begin{pmatrix} E_L \\ E_R \end{pmatrix} = \mathbf{WE}, \tag{5.2}$$

where we have dropped L and R, as external drive shifts the location of the equilibrium (where $E_L = E_R = 0$), but does not affect its stability (this is equivalent to studying the Jacobian matrix of equation 6.1). Winner-take-all behaviour requires that the equilibrium of the system is not asymptotically stable, that is, the matrix $\mathbf{W}$ must have at least one positive eigenvalue. The eigenvalues of $\mathbf{W}$ are: $\lambda_1 = \tau_E^{-1}(-1 - a + \varepsilon)$, and $\lambda_2 = \tau_E^{-1}(-1 + \varepsilon + a)$. If $a, \varepsilon > 0$, and $a > \varepsilon$, which is true for all dynamically plausible parameter values, then $\lambda_1 < 0$, leading to the following inequality for $\lambda_2 > 0$:

$$a > 1 - \varepsilon. \tag{5.3}$$

If the inequality in equation 6.3 is satisfied, unless the system starts in a precisely balanced initial condition with equal external drive (i.e. $E_L(0) = E_R(0)$ and $L = R$) the model will converge to a winner-take-all state (i.e. $E_L > 0$ and $E_R = 0$ or $E_L = 0$ and $E_R > 0$). Once in a winner-take-all state, as long as the competing population is suppressed, the state is stable.

Perceptual switches occur when the suppressed population escapes suppression. For example, if E_L is dominant, and E_R is suppressed, a perceptual switch occurs when E_R escapes from inhibition, which for the simple threshold nonlinearity occurs when the argument of $f(x)$ passes through zero from below (c.f. [47]). We, therefore, refer to the argument of the population transfer function as aggregate synaptic drive D which we define as a quantity in its own right and show below for the right monocular population:

$$D_R = I_R + \varepsilon E_R(t) - a E_L(t) - g_R H_R(t). \tag{5.4}$$

Aggregate synaptic drive takes negative values when the population is suppressed and positive values when dominant. We can find the minimum value of g_L necessary for the right monocular population to escape from suppression asymptotically by demanding that $D_R > 0$ and solving for g_L leading to:

$$g_L > a\frac{L}{R} - 1 + \varepsilon, \tag{5.5}$$

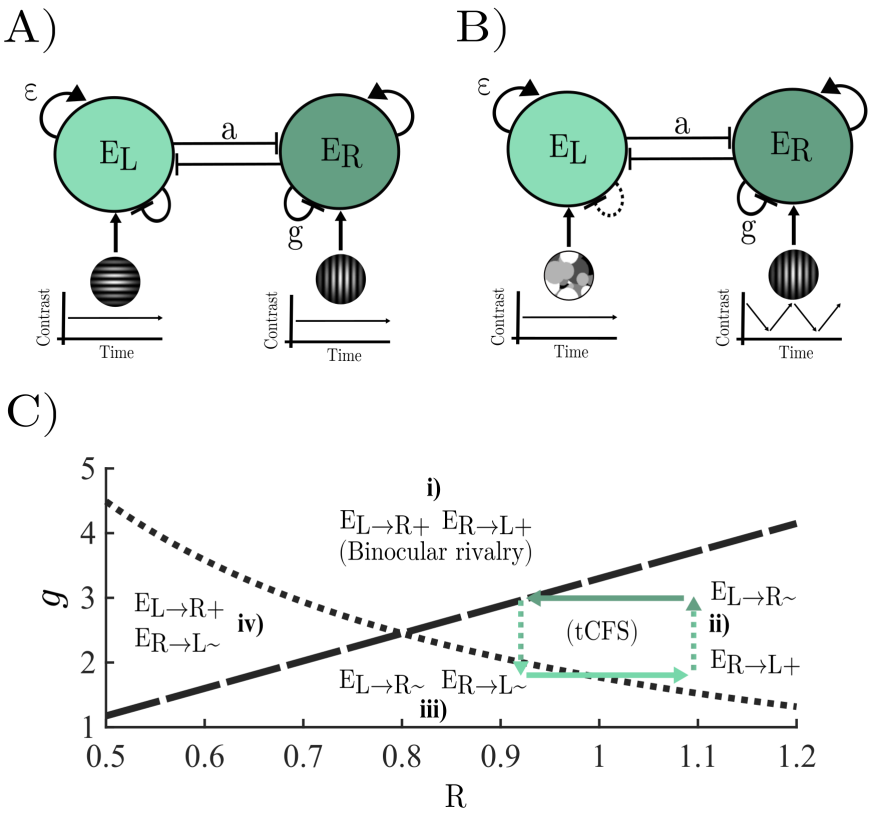

Figure 5.1: **A)** Model architecture for binocular rivalry two monocular populations compete for dominance through a process of mutual winner-take-all inhibition that is asymptotically destabilised by the accumulation of a slow hyperpolarising adaptation current. **B)** Model architecture for tCFS two slowly adapting monocular populations compete through mutual inhibition, as they do in binocular rivalry, but instead of being driven by two constant stimuli, the hypothesised effect of the dynamic mask is approximated by driving one population (E_L) with a stimulus of constant contrast and reducing the strength of adaptation, whilst driving the other population (E_R) with a dynamic stimulus (equation 6.6) that increases linearly when the population driven by the target stimulus is suppressed, and decreases linearly when it is dominant. See main text for a more in depth explanation. **C)** Adaptation by stimulus strength parameter space partitioned into four distinct dynamical regimes by the adaptation strength inequalities in equation 6.5 (g_L - dotted line, g_R - dashed line); regime i) both E_R and E_L are asymptotically released from inhibition by the accumulation of adaptation in the dominant population modelling typical binocular rivalry conditions; regime ii) The population driven by the mask (E_L) is asymptotically released from inhibition by the target stimulus population (E_R), but if the mask is dominant adaptation alone will not release the target stimulus population from inhibition; regime iii) both the mask and target stimulus population cannot escape from suppression by adaptation alone; regime iv) the population driven by the mask cannot escape suppression via adaptation alone, whilst the population driven by the target stimulus can escape suppression. Green arrows show a typical trajectory through the parameter space across a full dominance-suppression cycle in tCFS. The target stimulus starts close to full contrast making the strongly adapting target stimulus population dominant (dark green). Stimulus contrast then decreases until the weakly adapting mask driven population (light green) is released from suppression and takes over at which point the contrast of the stimulus starts to increase until the population driven by the stimulus is released from suppression.

where we have leveraged the fact that when the population is suppressed $\varepsilon E_R = 0$, and that at a perceptual switch the model is in the second asymptotic phase of its dynamics where $H_R \approx 0$, and $H_L \approx E_L$. We find the equilibrium value of the dominant population by substituting $H_L = E_L$ into the equation for $\dot{E}_L = 0$, and solve for E_L to obtain $E_L(\infty) = \frac{L}{1+g_L-\varepsilon}$ which, when substituted into equation 6.4, leads to the above inequality. The comparable expression

for the asymptotic release of E_L from suppression is given by $g_R > a\frac{R}{L} - 1 + \varepsilon$. This leaves us with three inequalities that when satisfied guarantee a limit cycle with dynamics that oscillate between mutually exclusive states of perceptual dominance.

To simulate binocular rivalry which presents stationary stimuli with matched image statistics to either eye we set L and R to constant values with equal adaptation strengths for each monocular population ($g_L = g_R$; Fig 5.1A). Under these conditions (in the absence of noise) the only way for a population to escape from inhibition, and a perceptual switch to occur, is for the model to satisfy the above system of inequalities guaranteeing a limit cycle. These inequalities define the minimal physiological conditions for binocular rivalry in terms of the parameters of the deterministic dynamical system given in equation 6.1.

Unlike binocular rivalry, where perceptual switches occur every few seconds, in tCFS, perceptual suppression can last for tens of seconds and switches are primarily determined by increasing/decreasing the contrast of the target stimulus. Indeed, in tCFS the stimuli entering both eyes are non-stationary. A dynamic (flashing) mask is presented to one eye and the contrast of the target stimulus presented to the other eye changes linearly with time. We approximated the effect of the dynamic mask (shown to E_L), which is hypothesised to reduce the accumulation of adaptation, by leaving the value of L fixed and reducing the value of g_L relative to g_R. We arrived at this approximation based upon two considerations. First, although the content of the mask is updated dynamically, the contrast of the mask is constant. Second, in a related modelling study, [58] showed that the dynamic flashing effect of the mask can be accurately modelled by periodically exciting neurons with different selectivity profiles effectively slowing the accumulation of adaptation. We settled on the precise values of g_L and L used in simulation by running a grid search over values of g_L and L to find the parameter combination that best minimised the difference between the simulated and empirical dominance durations of [38] across contrast rates. Full details are given in the next chapter. The optimal values ($g_L = 1.7$, $L = 0.8$) violate the inequality in equation 6.5 ($g_L < \frac{aL}{R} - 1 + \varepsilon$), for an adaptation driven release from inhibition (Fig 5.1B). This means that the population driven by the target stimulus (E_R) cannot escape suppression by means of adaptation alone.

In tCFS, as described above, the contrast of the target stimulus increases until the population is released from suppression, thereby gaining dominance, and then decreases until it is suppressed again. To model this process we used a piecewise linear function whose value was dependent on which monocular population was dominant:

$$R(t + \Delta t) = \begin{cases} R(t) - \gamma \Delta t & \text{if } E_R > E_L \\ R(t) + \gamma \Delta t & \text{if } E_R < E_L \end{cases} \tag{5.6}$$

where γ is a unitless rate constant. In Fig 5.1C we show a typical trajectory through the parameter space defined by $g_{L/R}$ and $R(t)$ for a single trial of tCFS. The parameter space is partitioned into four regions by the inequalities given by equation 6.5 which define the models dynamical regime. With the exception of a brief period following the suppression of the target stimulus, all values of R visited by the model allow the monocular population driven by the mask (E_L) to asymptotically escape from inhibition via the accumulation of adaptation and do not allow the population driven by the target stimulus (E_R) to escape from suppression via adaptation alone. That is, escaping from suppression (perceptual breakthrough) requires R to increase sufficiently so that $D_R > 0$, adaptation contributes to this process through a reduction in competitive inhibition from the dominant population, and a reduction in self-inhibition, but is insufficient to lead to a switch without an increase in R.

All simulations were run by integrating equation 6.1 numerically using the forward Euler (or Euler-Maruyama when noise was present) method with a time step of $dt = 0.1$ ms in MATLAB 2023b. Model parameters for each simulation are supplied in appendix 1.

5.3 Chapter 5 results

5.3.1 Suppression depth in binocular rivalry

When driven with constant input, the model oscillates between mutually exclusive states of perceptual dominance (Fig 5.2A) defined by the dominance of one monocular neural population and the suppression of the other (Fig 5.2A upper). Switches occur as the adaptation current approaches its asymptotic value (Fig 5.2A middle) and the aggregate synaptic drive of the non-dominant population passes through the rectification threshold at zero (Fig 5.2C lower), allowing the previously suppressed monocular population to inhibit its competitor and gain dominance. Inspection of the aggregate synaptic drive (Fig 5.2A lower) shows that it has a negative peak shortly after each perceptual switch and then decays exponentially until it passes through the rectification threshold and the suppressed population becomes dominant again.

This leads to a straightforward prediction which is consistent with prior psychological theory [1]: the depth of suppression for stimuli with similar selectivity profiles to the rivalry stimuli should decrease as a function of time relative to the last perceptual switch as inhibition from the dominant population, and self-adaptation, decrease. Contrary to this prediction, initial studies found that the detection threshold for stimuli presented to the suppressed eye was constant [61, 62]. However, the stimuli used in each study differed in both spatial frequency and orientation to the rivalry stimuli, making it unlikely that the populations excited by the probes were under inhibitory pressure from self-adaptation, or the competing monocular population. In addition, probes were presented in the first half of the suppression period following a perceptual switch where adaptation, which has a time constant on the order of a second [59, 60], is unlikely have had sufficient time to exert a large effect on population firing rates.

Indeed, in line with the prediction that suppression depth will decrease with time, [37] found that when probe stimuli matched the spatial frequency and orientation of the rivalry stimuli, and were aligned in time relative to the last perceptual switch, the probability of successful probe detection for stimuli presented to the suppressed eye increased with time relative to the last perceptual switch (Fig 5.2C dashed line). Similarly, the probability of successful probe detection for stimuli presented to the dominant eye (slightly) reduced with time (Fig 5.2C solid line).

To understand these results in more detail, beyond the simple qualitative prediction outlined above, we explicitly simulated a threshold detection extension of binocular rivalry. Following the experimental protocol, each simulation lasted 3 minutes with 60 probes presented to one monocular population at random times over the course of each trial (i.e. a probe was presented approximately every 3 seconds). To generate conditions closer to the *in vivo* experimental conditions, we added Gaussian noise ($\sigma = 0.0025$) to the adaptation current. For robustness, results were averaged over 40 simulations, each with independently generated noise. Each probe was presented for 10 ms of simulation time with strength ranging from [0.1 1]. We categorised each probe presentation according to whether the probe was presented in the

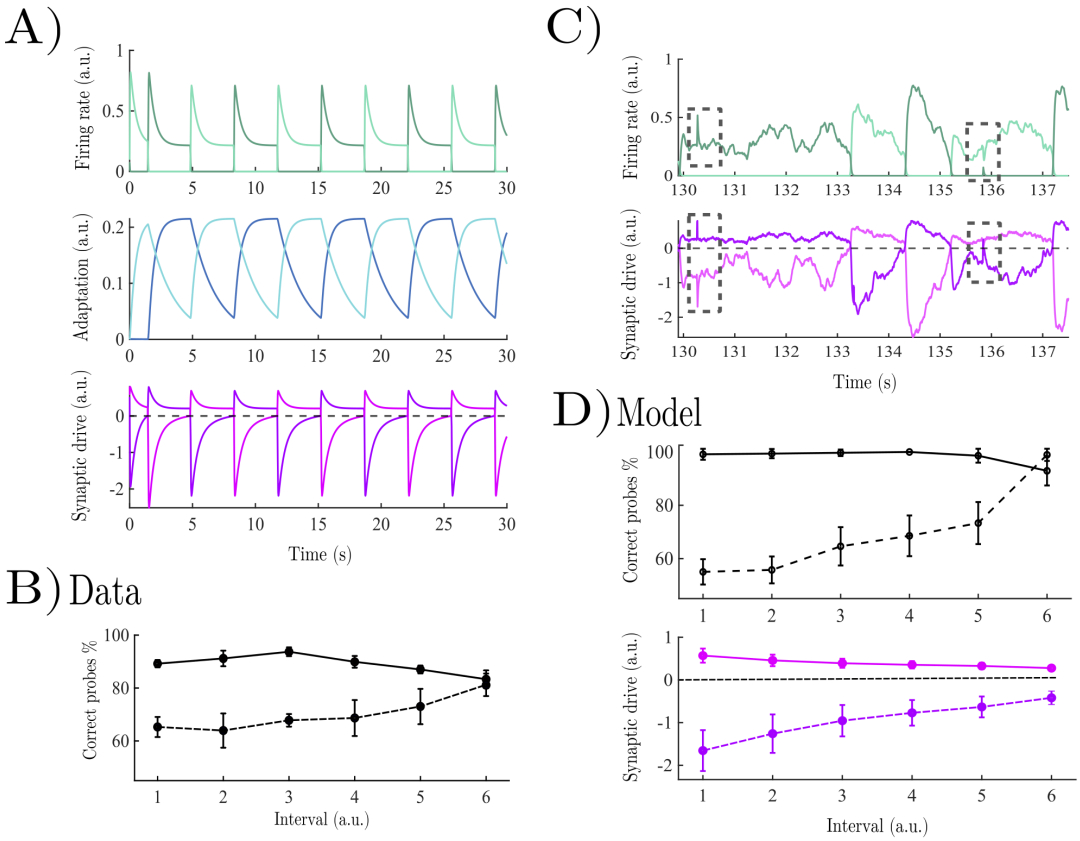

Figure 5.2: **A)** Model dynamics for a typical binocular rivalry simulation; upper) firing rate of each competing monocular population; middle) adaptation dynamics; lower) aggregate synaptic drive. **B)** Empirical data adapted from the threshold detection extension of binocular rivalry of [37]. Solid lines show the percentage of correct participant responses to probes presented to the dominant eye, dash lines show the percentage of correct participant responses to probes presented to the suppressed eye. **C)** Example firing rates and synaptic drive for competing populations receiving probes during periods of dominance and suppression. **D)** Simulated model behaviour; upper) solid lines show the percentage of correct ideal observer responses to probes presented to the dominant eye, dashed lines show the percentage of correct ideal observer responses to probes presented to the suppressed eye; lower) aggregate synaptic drive for the dominant (solid pink line) population, and suppressed (dashed purple line) populations in each normalised time interval.

dominant or suppressed period, and sorted neural activity into 6 time bins normalised by the relative length of each suppression/dominance period (as was done empirically). We modelled behavioural responses in terms of what an ideal observer with an optimal criterion (i.e., a criterion selected to minimise misses and false alarms across all six normalised time bins) could readout from the summed population firing rate during the probe presentation window, with independent criteria for dominant and suppressed periods (for details see Appendix 2). As with the empirical paradigm, probes that induced perceptual switches were excluded from further analysis. For simplicity, in the main text we only show the simulation results for probe strength values of 0.9 but note that the results hold qualitatively across the full range of probe strengths (see Appendix 2).

In Fig 5.2C, we show an example of the model dynamics, as a probe is presented to one monocular population in a period of dominance and then suppression. Probes presented whilst the population was dominant generated a spike in the firing rate of the dominant population that could be reliably readout by the ideal observer with occasional misses caused by the noisy

adaptation current (solid line Fig 5.2D upper). Probes presented to the suppressed population broke though less reliably. In Fig 5.2C, the probe coincides with a positive fluctuation in aggregate synaptic drive allowing the aggregate synaptic drive to transiently pass through the rectification threshold and generate a brief spike in the firing rate of the suppressed population. The probability of such breakthroughs occurring increased as a function of time relative to the last perceptual switch (Fig 5.2D upper). Indeed, the behavioural response curves are essentially a mirror image of the aggregate synaptic drive in each normalised interval flipped about the x-axis (Fig 5.2D lower). Here, therefore, in line with empirical data and previous psychological theory [1, 37], the depth of perceptual suppression in the model decreased as a function of time relative to the last perceptual switch. This is due to the reduction in self-inhibition in the suppressed monocular population, and the reduction in inhibition from the dominant population, both of which were due to the adaptation dynamics pushing the aggregate synaptic drive toward the rectification threshold at zero.

5.3.2 Suppression depth in tracking continuous flash suppression

The difference in the probability of successful probe detection between the suppressed and dominant eyes in binocular rivalry serves as proxy for the depth of perceptual suppression, an inference supported by the simulations in the previous section, where the probability of successful detection mirrors the aggregate synaptic drive, D. Although having the advantage of being time resolved, this experimental approach does not provide an absolute measure of suppression depth. In this section, we generalise the model to tCFS (i.e., tracking continuous flash suppression), a paradigm that provides a stationary, but absolute, measure of suppression depth by measuring the threshold for both breakthrough and suppression (Fig 5.3A; analogous to measuring the difference between the threshold for the successful detection of probes presented to the dominant and suppressed eyes).

In tCFS, perceptual switches are governed by gradually increasing/decreasing the strength of one monocular stimulus, whilst presenting a high contrast mask to the other eye (Fig 5.1B). The target stimulus starts at full contrast and then linearly decreases (in decibel scale; dB) until it vanishes from awareness and is suppressed by the mask (indexed by a button press), at which point the stimulus increases in contrast until it escapes suppression from the mask and breaks through into awareness (again indexed by a button press) and the cycle repeats [38]. Intriguingly, although the threshold for breakthrough varies across stimulus types the difference between breakthrough and suppression thresholds is constant. The threshold for breakthrough is always larger than the threshold for suppression (Fig 5.3A), suggesting a shared hysteretic mechanism of perceptual breakthrough and suppression that occurs early in visual processing. In addition, the rate of contrast change modulates the depth of perceptual suppression, with faster rates increasing suppression depth ([38]; Fig 5.3D upper) suggestive of an adaptation driven reduction in suppression depth consistent with the threshold detection binocular rivalry results explored in the previous section [37] and the subtractive effect of adaptation in the model.

To formalise and test this hypothesis *in silico*, we simulated tCFS across a range of contrast rates within our minimal model. Importantly, because the mechanism of suppression in the model relies upon competitive inhibition between monocular neural populations, the model is a good representation of the hypothesis that the process of interocular suppression and breakthrough occurs between ocular dominance columns in early visual cortex. Following the experimental protocol of [38], we simulated tCFS by replacing the constant stimulus driving

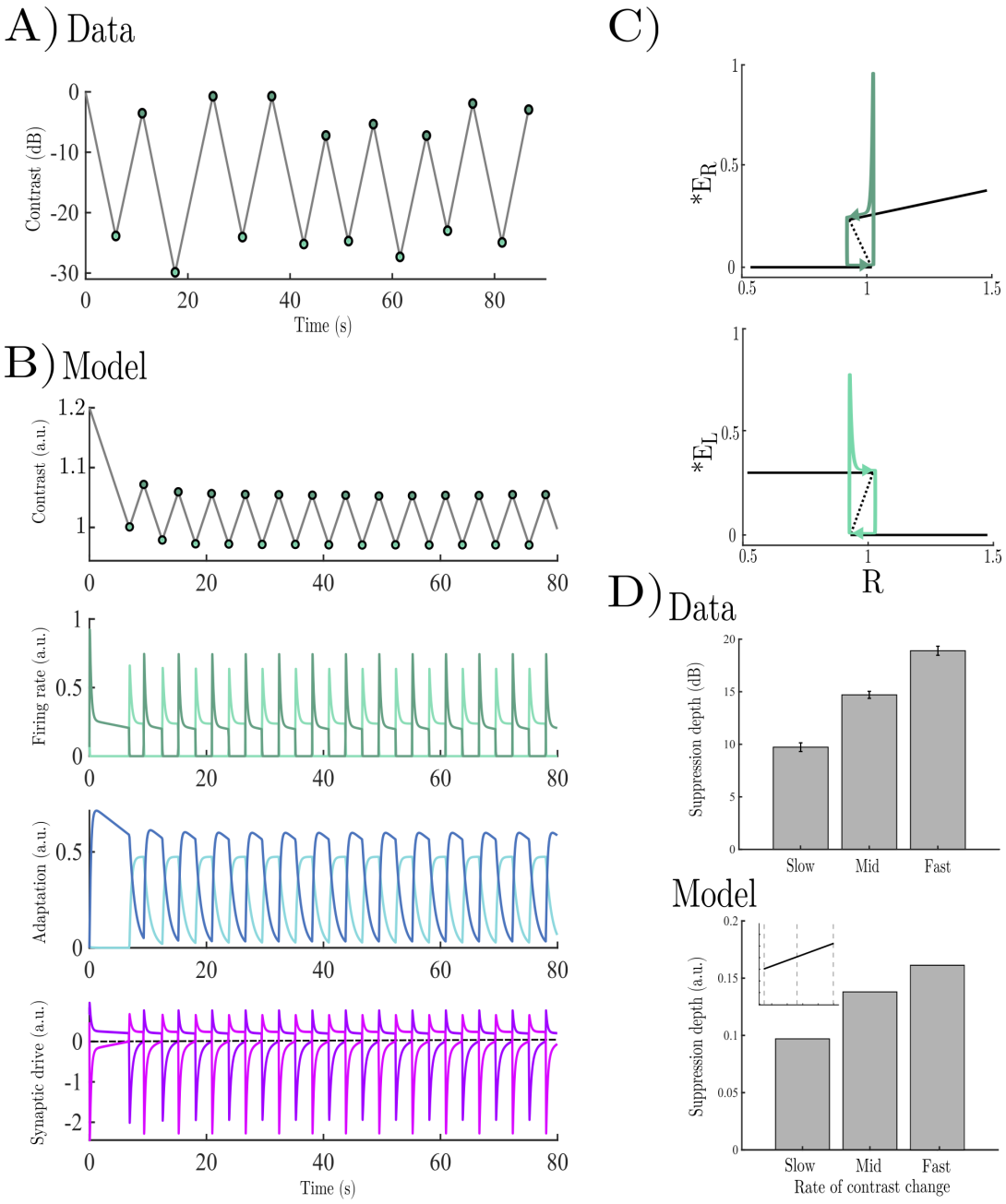

Figure 5.3: **A)** Empirical stimulus contrast dynamics from an example tCFS trial adapted from [38]. Light green dots show points of perceptual suppression, and dark green dots show points of perceptual breakthrough. **B)** Model dynamics for a typical tCFS simulation; upper) stimulus contrast dynamics; upper middle) firing rate of each competing monocular population; lower middle) adaptation dynamics; lower) aggregate synaptic drive. **C)** Bifurcation diagram for typical firing rate trajectories across the dominance-suppression cycle. Solid lines show the location of stable equilibrium values of $E_R = 0$ and $E_L = 0$ when dominant and suppressed. Dashed line shows unstable equilibrium location. **D)** Suppression depth as a function of contrast rate; upper) empirical data adapted from [38]; lower) simulated data from model, inset shows the three equally spaced values represented bar graph across all simulated contrast rate simulations.

the right monocular population with a piecewise linear function (equation 6.6) that linearly decreased/increased the contrast of the target stimulus depending on which monocular population was dominant. We approximated the effect of the dynamic mask by driving the other monocular population with a constant stimulus and reducing the strength of adaptation below

the threshold for an adaptation-driven release from inhibition based on the hypothesis that refreshing the content of the stimulus reduces the accumulation of adaptation by stimulating neurons with distinct selectivity profiles [38].

All simulations started with a high contrast target stimulus, which decreased in contrast until the population driven by the mask escaped suppression gaining dominance by inhibiting the stimulus-driven population into silence, at which point the contrast of the target stimulus increased again, and the cycle repeated (Fig 5.3B). As with the empirical data, the model converged to an equilibrium with stable breakthrough and suppression values within a few contrast reversal cycles post-trial onset, showing the same hysteretic pattern of higher stimulus contrast for breakthrough thresholds than suppression thresholds. This hysteretic cycle of breakthrough and suppression is shown in the bifurcation diagrams in Fig 5.3C. Upon breaking free from suppression, each population jumps past the upper-stable branch of the bifurcation diagram and then exponentially approaches the stable steady state given by the expressions for $E_R(\infty)$ and $E_R(\infty)$. The population driven by the (decreasing) target stimulus is further from the stable steady state than the mask-driven population as the contrast of the stimulus decreases faster than the intrinsic time scale of the system. Instead, the system chases (but never reaches) the stable fixed point at $E_R(t) = (R(t) - \gamma t)/(1 + g_R - \varepsilon)$ (c.f. [63]). Perceptual switches occur when the stimulus decreases sufficiently to release the mask from suppression (i.e., when $D_L = 0$) which then, in turn, suppresses the target stimulus population and the stimulus increases in contrast until $D_R = 0$.

The magnitude of this hysteretic effect (i.e., the difference between the breakthrough and suppression thresholds) is what is quantified by experimental measures of suppression depth. Explaining the determinants of hysteresis in the model, therefore, amounts to explaining the determinants of suppression depth. The key feature of the model responsible for the presence of hysteresis in the dynamics is the strong recurrent inhibition. This can be seen by deriving approximate expressions for the strength of the target stimulus R at breakthrough (R_B) and suppression (R_S) points. At a perceptual breakthrough $E_L(t) \approx E_L(\infty) = \frac{L}{1+g_L-\varepsilon}$, $E_R = 0$, and $D_R = 0$. Substituting these values into equation 6.4 and solving for $I_R = R_B$ we obtain:

$$R_B = \frac{aL}{1 + g_L - \varepsilon} + g_R H_R(t). \tag{5.7}$$

Analogously, when the population driven by the target stimulus is suppressed $E_R(t) \approx E_R(\infty) = \frac{R_S}{1+g_R-\varepsilon}$, $E_L = 0$, and $D_L = 0$, leading to:

$$R_S = L\frac{(1 + g_R - \varepsilon)}{a} - g_L H_L(t)\frac{(1 + g_R - \varepsilon)}{a}. \tag{5.8}$$

Focusing on the constant terms, we see that competitive inhibition (a) is in the numerator of R_B and in the denominator of R_S. This implies that strong competitive inhibition increases the stimulus contrast needed for a population to escape from suppression and decreases the stimulus contrast needed to inhibit the competing population into silence. Whereas increasing the strength of adaptation for the population driven by the target stimulus has the opposite effect. When the contrast accumulation rate (γ) is slow, the contribution of the time-dependent adaptation terms is negligible (i.e., $H_R, H_L \approx 0$), and the magnitude of suppression depth is entirely driven by strong competitive inhibition. As the rate of contrast accumulation increases, adaptation in the suppressed population has less time to decay, increasing suppression depth through the time-dependent terms in equations 6.7-6.8, which have an additive effect on the breakthrough threshold (R_B) and a subtractive effect on the suppression threshold

(R_S). This leads to a net increase in suppression depth with increased contrast rate as is seen empirically (Fig 5.3D). This makes intuitive sense; the subtractive effect of adaptation means that higher external drive from the target stimulus is required to break free from suppression, and likewise, less inhibition is required to keep another population suppressed.

5.4 Chapter 5 discussion

The central appeal of visual rivalry in the experimental study of conscious perception is that two physically-matched stimuli enter the visual system, but only one is consciously experienced, whilst the other is suppressed from awareness. This provides a well-controlled experimental paradigm in which to measure, and potentially mechanistically explain, the threshold for visual awareness, and the suppression of content from awareness.

In stark contrast to standard binocular rivalry, which has benefited from quantitative neural theory for over half a century, extensions of the standard binocular rivalry paradigm that allow researchers to quantify the depth of perceptual suppression have received little to no attention with the notable exception of [58] who modified Wilson's [49] minimal model to explain the protracted suppression times known to characterise continuous flash suppression (CFS). Building on this work, we extended the same [49] model to encompass two empirical paradigms that provide complementary insights into the dynamics [37] and magnitude [38] of suppression depth in visual rivalry.

We used the minimal model of Wilson [49] as our foundation as it lives in the ideal zone of complexity with just enough machinery to account for the key phenomena of interest, whilst still being simple enough that it is possible to derive insights from approximations to the dynamics of the full system. In addition, the dynamics of the model are governed by two key qualitative features, strong competitive inhibition and self-adaptation, common to almost all neuronal models of rivalry [47]. This commonality means that the inferences made in the paper are granted a greater degree of generality than would be possible if we had used a more idiosyncratic model as our starting point.

In the first simulation we showed, in line with empirical data [37], that the depth of perceptual suppression in binocular rivalry decreases as a function of time relative to the last perceptual switch. The model shows that this is due to the reduction in self-inhibition in the suppressed monocular population as the population recovers from adaptation accumulated in the preceding dominance period, and a reduction in inhibition from the dominant population as adaptation accumulates. Together, this pushes the aggregate synaptic drive in the suppressed population toward the rectification threshold at zero, increasing the probability with which probe stimuli presented to the suppressed eye will be able to transiently break free from suppression.

In the second simulation, we showed - in line with the predictions of [64] - that the depth of perceptual suppression in tracking continuous flash suppression (tCFS) can be explained by a combination of competitive inhibition and adaptation with monocular populations reminiscent of ocular dominance columns in early visual cortex. When the rate of contrast change is slow, adaptation in the suppressed monocular population has a large amount of time to decay. In this case, the hysteretic difference in thresholds for breakthrough and suppression is due solely to competitive inhibition increasing the stimulus contrast needed for a population to escape suppression, and correspondingly, reducing the contrast needed to inhibit the competing population into suppression. As the contrast accumulation rate increases adaptation

in the suppressed population has less time to decay and adaptation accumulated in previous periods of dominance carries over across periods of suppression. Crucially, the subtractive effect of adaptation means that suppression depth increases as a function of the rate of contrast change as more external drive from the stimulus is needed in order for the population to break free from suppression, and correspondingly, less inhibition is required to keep the competing population suppressed.

The location of competition in the model most plausibly corresponds to competition between ocular dominance columns within early visual cortex that are biased towards input from one eye [44, 49]. One possible criticism, therefore, is that the model does not capture findings showing that rivalry can occur at multiple stages in the visual hierarchy depending on stimulus type (e.g. [65, 66]). This is an excellent point. We are not claiming that rivalry necessarily occur within early visual cortex, rather in line with [49], we argue that competitive inhibition between V1-like monocular populations is physiologically sufficient to explain the modulation of suppression depth in standard binocular rivalry and tCFS, a claim that is entirely compatible with multilevel models of more complicated varieties of binocular rivalry which allow competition to occur at a later stage of hierarchical processing [44, 67, 68]. Indeed, our monocular focus is also supported by the current state of the art in non-human primate neurophysiology. In a recent study [69] using two-photon calcium imaging it was shown that when monocular-biased neurons in V1 are suppressed by a mask in CFS their orientation tuning amplitude decrease by $\sim 80\%$.

Another potential criticism is that the model does not capture the finding that suppression depth is constant across stimulus types. Given the simplicity of the model, this is necessarily the case. The model has no spatial extent and, therefore, cannot explain variation in stimulus content. This limitation in scope is, however, common across almost all computational models of binocular rivalry and is a result of the inherent trade-off between conceptual tractability and model complexity. As our goal was to explore the minimally sufficient physiological conditions for suppression depth across binocular rivalry and tCFS, we leaned heavily on the side of simplicity.

The key contribution of this paper is the unification of threshold detection variants of binocular rivalry and tCFS with one minimally sufficient mechanism. Suppression depth, in both paradigms, can be seen as a result of the same underlying competitive process between slowly adapting monocular populations. In binocular rivalry, the accumulation of adaptation in the dominant population, and the recovery from adaptation in the suppressed population, are the key factors driving the reduction of suppression depth with time. The increase in suppression depth as a function of contrast rate in CFS is a manifestation of this same underlying mechanism. As the contrast rate increases the (subtractive) contribution of adaptation to the aggregate synaptic drive is increased, meaning that a larger increase in stimulus contrast is needed for the population to escape inhibition. By the same token, once a population is suppressed, adaptation carried over from the previous dominance period means that less inhibition is required to keep the population suppressed.

Having established the face validity of our model of suppression depth through numerical simulations, in the next chapter, we will use the same model to generate closed-form expressions for the quantities qualitatively characterised here, and move beyond existing experimental data to generate novel empirical predictions based on the model.

5.5 Chapter 5 supplementary material

5.5.1 Model parameters

Parameter	Description	Value (BR)	Value (tCFS)	Units
τ_E	Neuronal population time constant	15	15	ms
τ_H	Hyperpolarizing adaptation current time constant	1950	1000	ms
L	External drive to left monocular population	0.85	0.8	a.u.
R	External drive to right monocular population	0.85	$R_0 = 1.2$	a.u.
ε	Strength of excitatory recurrent projections	0.05	0.05	a.u.
a	Strength of competitive inhibition	3.4	3.4	a.u.
g_L	Strength of left population adaptation current	3	1.7	a.u.
g_R	Strength of right population adaptation current	3	3	a.u.
γ	Rate of contrast change	n/a	$(0.0021 - 0.0063) \times 10^{-2}$	a.u.

Table 5.1: Parameter description, values, and units of the model described by equations 6.1-6.6. A longer adaptation time constant was used for the binocular rivalry (BR) simulations to generate empirically realistic dominance durations in the presence of noise.

5.5.2 Ideal observer readout

To obtain behavioural responses from the model that are comparable to the participant responses in [37] we constructed an ideal observer based upon the summed evoked firing rate of the monocular population receiving the probe stimulus. We sorted each probe presentation according to whether they occurred during dominance or suppression periods and sorted neural activity into 6 time bins with normalised durations defined by the length of each suppression and dominance period. As in the empirical analysis, probe presentations that led to a perceptual switch were excluded from further analysis. In addition, we excluded probe presentaions that occurred during dominance and suppressin periods less than 600 ms in duration so that the minimium normalised time window was at least 100 ms.

Across probe contrast levels, including probe-absent trials, we calculated the frequency with which the summed firing rate in the 10 ms probe stimulus presentation window exceeded a criterion (i.e., $\sum_{t=\text{stim on}}^{T=\text{stim off}} E_R(t) >$ criterion). The criterion was defined on the interval between the minium and maximium summed firing rate across all probe strengths. We then selected the (optimal) criterion that best minimised misses and false alarms across all six of the duration-normalised time bins. Criterions for suppressed and dominant populations were calculated independently. Summed firing rates exceeding the optimal criterion were counted as a behavioural response. As the model was not forced to emitt a response we normalised the model response probability (on the inverval [0 1]) to vary between [0.5 1], under the assumption that in the absence of any information participants (and the model) make an unbiased guess. Model responses across the full range of probe contrast values are shown in Fig 5.5.2.

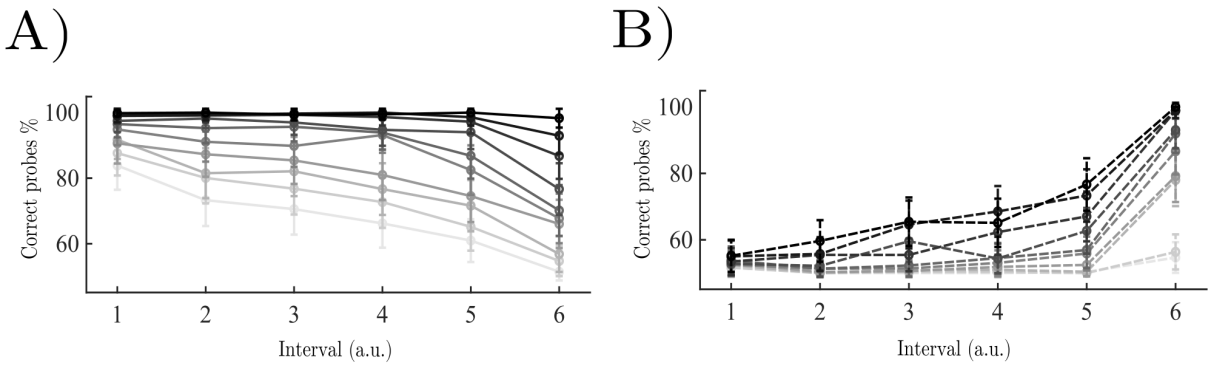

Figure 5.4: Simulated model behaviour across the full range of probe stimulus contrast values from 0.1 (lightest grey) to 1 (black). **A)** percentage of correct ideal observer responses to probes presented to the dominant monocular population. **B)** Same but for the suppressed population.

Chapter 5 references

[1] Randolph Blake. "A neural theory of binocular rivalry". In: *Psychological Review* 96.1 (1989), pp. 145–167. DOI: 10.1037/0033-295X.96.1.145.

[2] Randolph Blake and K. Boothroyd. "The precedence of binocular fusion over binocular rivalry". In: *Perception & Psychophysics* 37.2 (1985), pp. 114–124. DOI: 10.3758/BF03202845.

[3] Randolph Blake. "A Primer on Binocular Rivalry, Including Current Controversies". In: *Brain and Mind* 2.1 (2001), pp. 5–38. DOI: 10.1023/A:1017925416289.

[4] R. Fox and J. Herrmann. "Stochastic properties of binocular rivalry alternations". In: *Perception & Psychophysics* 2.9 (1967), pp. 432–436. DOI: 10.3758/BF03208783.

[5] Randolph Blake and R. Fox. "Binocular rivalry suppression: Insensitive to spatial frequency and orientation change". In: *Vision Research* 14.8 (1974), pp. 687–692. DOI: 10.1016/0042-6989(74)90065-0.

[6] W.J.M. Levelt. "On binocular rivalry". PhD thesis. Soesterberg, The Netherlands: Institute for Perception RVO-TNO, 1965.

[7] B. B. Breese. "On inhibition". In: *The Psychological Review: Monograph Supplements* 3.1 (1899), pp. i–65. DOI: 10.1037/h0092990.

[8] R. S. Creed. "Observations on binocular fusion and rivalry". In: *The Journal of Physiology* 84.4 (1935), pp. 381–392. DOI: 10.1113/jphysiol.1935.sp003288.

[9] David Alais. "Binocular rivalry: Competition and inhibition in visual perception". In: *WIREs Cognitive Science* 3.1 (2012), pp. 87–103. DOI: 10.1002/wcs.151.

[10] Randolph Blake and Nikos K. Logothetis. "Visual competition". In: *Nature Reviews Neuroscience* 3.1 (2002), pp. 13–21. DOI: 10.1038/nrn701.

[11] Jan Brascamp, Randolph Blake, and Tomas Knapen. "Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception". In: *Nature Neuroscience* 18.11 (2015), pp. 1672–1678. DOI: 10.1038/nn.4130.

[12] J.-D. Haynes, R. Deichmann, and G. Rees. "Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus". In: *Nature* 438.7067 (2005), pp. 496–499. DOI: 10.1038/nature04169.

[13] E. D. Lumer, K. J. Friston, and G. Rees. "Neural Correlates of Perceptual Rivalry in the Human Brain". In: *Science* 280.5371 (1998), pp. 1930–1934. DOI: 10.1126/science.280.5371.1930.

[14] A. Polonsky et al. "Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry". In: *Nature Neuroscience* 3.11 (2000), pp. 1153–1159. DOI: 10.1038/80676.

[15] F. Tong et al. "Binocular Rivalry and Visual Awareness in Human Extrastriate Cortex". In: *Neuron* 21.4 (1998), pp. 753–759. DOI: 10.1016/S0896-6273(00)80592-9.

[16] G. Tononi and G.M. Edelman. "Consciousness and Complexity". In: *Science* 282 (1998), pp. 1846–1851.

[17] A. Dwarakanath et al. "Prefrontal state fluctuations control access to consciousness". In: *bioRxiv* (2020). [Preprint]. DOI: 10.1101/2020.01.29.924928.

[18] J. K. Hesse and D. Y. Tsao. "A new no-report paradigm reveals that face cells encode both consciously perceived and suppressed stimuli". In: *eLife* 9 (2020), e58360. DOI: 10.7554/eLife.58360.

[19] V. Kapoor et al. "Decoding the contents of consciousness from prefrontal ensembles". In: *bioRxiv* (2020). [Preprint]. DOI: 10.1101/2020.01.28.921841.

[20] N. K. Logothetis and D. A. Leopold. "What is rivalling during binocular rivalry?" In: *Nature* 380 (1996), pp. 621–624.

[21] N. K. Logothetis. "Single units and conscious vision". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 353.1377 (1998), pp. 1801–1818.

[22] T. I. Panagiotaropoulos et al. "Neuronal Discharges and Gamma Oscillations Explicitly Reflect Visual Consciousness in the Lateral Prefrontal Cortex". In: *Neuron* 74.5 (2012), pp. 924–935. DOI: 10.1016/j.neuron.2012.04.013.

[23] M. Wilke, K.-M. Mueller, and D.A. Leopold. "Neural activity in the visual thalamus reflects perceptual suppression". In: *Proceedings of the National Academy of Sciences of the United States of America* 106 (2009), pp. 9465–9470.

[24] Francis Crick. *Astonishing Hypothesis: The Scientific Search for the Soul*. Simon and Schuster, 1995.

[25] David Alais and David Melcher. "Strength and coherence of binocular rivalry depends on shared stimulus complexity". In: *Vision Research* 47.2 (2007), pp. 269–279. DOI: 10.1016/j.visres.2006.09.003.

[26] D. Apthorp, P. Wenderoth, and D. Alais. "Motion streaks in fast motion rivalry cause orientation-selective suppression". In: *Journal of Vision* 9.5 (2009), p. 10. DOI: 10.1167/9.5.10.

[27] Randolph Blake and J. Camisa. "On the inhibitory nature of binocular rivalry suppression". In: *Journal of Experimental Psychology: Human Perception and Performance* 5.2 (1979), pp. 315–323. DOI: 10.1037/0096-1523.5.2.315.

[28] M. Hollins and G. Bailey. "Rivalry target luminance does not affect suppression depth". In: *Perception & Psychophysics* 30 (1981), pp. 201–203. DOI: 10.3758/BF03204480.

[29] C. Lunghi and D. Alais. "Congruent tactile stimulation reduces the strength of visual suppression during binocular rivalry". In: *Scientific Reports* 5.1 (2015), p. 9413. DOI: 10.1038/srep09413.

[30] V. A. Nguyen, A. W. Freeman, and D. Alais. "Increasing depth of binocular rivalry suppression along two visual pathways". In: *Vision Research* 43.19 (2003), pp. 2003–2008. DOI: 10.1016/S0042-6989(03)00314-6.

[31] O. Teng Leng and M. S. Loop. "Visual suppression and its effect upon color and luminance sensitivity". In: *Vision Research* 34.22 (1994), pp. 2997–3003. DOI: 10.1016/0042-6989(94)90272-0.

[32] N. Tsuchiya et al. "Depth of interocular suppression associated with continuous flash suppression, flash suppression, and binocular rivalry". In: *Journal of Vision* 6.10 (2006), p. 6. DOI: 10.1167/6.10.6.

[33] Y. Watanabe and S. Funahashi. "Neuronal Activity Throughout the Primate Mediodorsal Nucleus of the Thalamus During Oculomotor Delayed-Responses. I. Cue-, Delay-, and Response-Period Activity". In: *Journal of Neurophysiology* 92 (2004), pp. 1738–1755.

[34] N. Tsuchiya and C. Koch. "Continuous flash suppression reduces negative afterimages". In: *Nature Neuroscience* 8.8 (2005), pp. 1096–1101. DOI: 10.1038/nn1500.

[35] S. Gayet, S. Van der Stigchel, and C. L. E. Paffen. "Seeing is believing: Utilization of subliminal symbols requires a visible relevant context". In: *Attention, Perception, & Psychophysics* 76.2 (2014), pp. 489–507. DOI: 10.3758/s13414-013-0580-4.

[36] A. Pournaghdali and B. L. Schwartz. "Continuous flash suppression: Known and unknowns". In: *Psychonomic Bulletin & Review* 27.6 (2020), pp. 1071–1103. DOI: 10.3758/s13423-020-01771-2.

[37] David Alais et al. “Visual Sensitivity Underlying Changes in Visual Consciousness”. In: *Current Biology* 20.15 (2010), pp. 1362–1367. DOI: 10.1016/j.cub.2010.06.015.

[38] David Alais et al. “tCFS: A new CFS tracking paradigm reveals uniform suppression depth regardless of target complexity or salience”. In: *eLife* (2023). [Preprint]. DOI: 10.7554/eLife.91019.1.

[39] W. J. M. Levelt. “Note on the Distribution of Dominance Times in Binocular Rivalry”. In: *British Journal of Psychology* 58.1-2 (1967), pp. 143–145. DOI: 10.1111/j.2044-8295.1967.tb01068.x.

[40] Jan W. Brascamp, P. Christiaan Klink, and Willem J. M. Levelt. “The laws of binocular rivalry: 50 years of Levelts propositions”. In: *Vision Research* 109 (2015), pp. 20–37. DOI: 10.1016/j.visres.2015.02.019.

[41] S. R. Lehky. “An Astable Multivibrator Model of Binocular Rivalry”. In: *Perception* 17.2 (1988), pp. 215–228. DOI: 10.1068/p170215.

[42] Carlo R. Laing and Carson C. Chow. “A Spiking Neuron Model for Binocular Rivalry”. In: *Journal of Computational Neuroscience* 12.1 (2002), pp. 39–53. DOI: 10.1023/A:1014942129705.

[43] Z. Wang, W. Dai, and D. W. McLaughlin. “Ring models of binocular rivalry and fusion”. In: *Journal of Computational Neuroscience* 48.2 (2020), pp. 193–211. DOI: 10.1007/s10827-020-00744-7.

[44] Hugh R. Wilson. “Computational evidence for a rivalry hierarchy in vision”. In: *Proceedings of the National Academy of Sciences* 100.24 (2003), pp. 14499–14503. DOI: 10.1073/pnas.2333622100.

[45] Carlo R. Laing, Thomas Frewen, and Ioannis G. Kevrekidis. “Reduced models for binocular rivalry”. In: *Journal of Computational Neuroscience* 28.3 (2010), pp. 459–476. DOI: 10.1007/s10827-010-0227-6.

[46] R. Moreno-Bote, J. Rinzel, and N. Rubin. “Noise-Induced Alternations in an Attractor Network Model of Perceptual Bistability”. In: *Journal of Neurophysiology* 98.3 (2007), pp. 1125–1139. DOI: 10.1152/jn.00116.2007.

[47] A. Shpiro et al. “Dynamical Characteristics Common to Neuronal Competition Models”. In: *Journal of Neurophysiology* 97.1 (2007), pp. 462–473. DOI: 10.1152/jn.00604.2006.

[48] A. Shpiro et al. “Balance between noise and adaptation in competition models of perceptual bistability”. In: *Journal of Computational Neuroscience* 27.1 (2009), pp. 37–54. DOI: 10.1007/s10827-008-0125-3.

[49] Hugh R. Wilson. “Minimal physiological conditions for binocular rivalry and rivalry memory”. In: *Vision Research* 47.21 (2007), pp. 2741–2750. DOI: 10.1016/j.visres.2007.07.007.

[50] Hugh R. Wilson. “Binocular contrast, stereopsis, and rivalry: Toward a dynamical synthesis”. In: *Vision Research* 140 (2017), pp. 89–95. DOI: 10.1016/j.visres.2017.07.016.

[51] Peter Dayan. “A Hierarchical Model of Binocular Rivalry”. In: *Neural Computation* 10.5 (1998), pp. 1119–1135. DOI: 10.1162/089976698300017377.

[52] S. Gershman, E. Vul, and J. Tenenbaum. “Perceptual Multistability as Markov Chain Monte Carlo Inference”. In: *Advances in Neural Information Processing Systems*. Vol. 22. 2009.

[53] S. J. Gershman, E. Vul, and J. B. Tenenbaum. “Multistability and Perceptual Inference”. In: *Neural Computation* 24.1 (2012), pp. 1–24. DOI: 10.1162/NECO_a_00226.

[54] J. Hohwy, A. Roepstorff, and K. Friston. "Predictive coding explains binocular rivalry: An epistemological review". In: *Cognition* 108.3 (2008), pp. 687–701. DOI: 10.1016/j.cognition.2008.05.010.

[55] T. Parr et al. "Perceptual awareness and active inference". In: *Neuroscience of Consciousness* 2019.1 (2019), niz012. DOI: 10.1093/nc/niz012.

[56] S. Safavi and P. Dayan. "Multistability, perceptual value, and internal foraging". In: *Neuron* 110.18 (2022), 3020–3034.e6. DOI: 10.1016/j.neuron.2022.07.024.

[57] S. Safavi and P. Dayan. "A decision-theoretic model of multistability: Perceptual switches as internal actions". In: *bioRxiv* (2024). DOI: 10.1101/2024.12.06.627286.

[58] D. Shimaoka and K. Kaneko. "Dynamical systems modeling of Continuous Flash Suppression". In: *Vision Research* 51.6 (2011), pp. 521–528. DOI: 10.1016/j.visres.2011.01.009.

[59] D.A. McCormick. "Cholinergic and Noradrenergic Modulation of Thalamocortical Processing". In: *Trends in Neurosciences* 12 (1989), pp. 215–221.

[60] H. R. Wilson and J. D. Cowan. "Evolution of the Wilson-Cowan equations". In: *Biological Cybernetics* 115.6 (2021), pp. 643–653. DOI: 10.1007/s00422-021-00912-7.

[61] R. Fox and R. Check. "Independence between binocular rivalry suppression duration and magnitude of suppression". In: *Journal of Experimental Psychology* 93.2 (1972), pp. 283–289. DOI: 10.1037/h0032455.

[62] H. F. Norman, J. F. Norman, and J. Bilotta. "The temporal course of suppression during binocular rivalry". In: *Perception* 29.7 (2000), pp. 831–841. DOI: 10.1068/p3085.

[63] R. D. Beer. "Codimension-2 parameter space structure of continuous-time recurrent neural networks". In: *Biological Cybernetics* 116.4 (2022), pp. 501–515. DOI: 10.1007/s00422-022-00938-5.

[64] David Alais et al. "tCFS: A new CFS tracking paradigm reveals uniform suppression depth regardless of target complexity or salience". In: *eLife* 12 (2024), e91019.

[65] S.-H. Lee and R. Blake. "Rival ideas about binocular rivalry". In: *Vision Research* 39.8 (1999), pp. 1447–1454. DOI: 10.1016/S0042-6989(98)00269-7.

[66] N. K. Logothetis, D. A. Leopold, and D. L. Sheinberg. "What is rivalling during binocular rivalry?" In: *Nature* 380.6575 (1996), pp. 621–624. DOI: 10.1038/380621a0.

[67] A. W. Freeman. "Multistage Model for Binocular Rivalry". In: *Journal of Neurophysiology* 94.6 (2005), pp. 4412–4420. DOI: 10.1152/jn.00557.2005.

[68] H.-H. Li et al. "Attention model of binocular rivalry". In: *Proceedings of the National Academy of Sciences* 114.30 (2017). DOI: 10.1073/pnas.1620475114.

[69] C.-X. Chen et al. "Continuous flashing suppression of V1 responses and the perceptual consequences revealed via two-photon calcium imaging and transformer modeling". In: *bioRxiv* (2025), p. 2025.06.13.658791. DOI: 10.1101/2025.06.13.658791.

Chapter 6

A Minimal Quantitative Model of Perceptual Suppression and Breakthrough in Visual Rivalry

> ...one thing that I like about this example is that even though it is comparatively simple, it exposes an important truth about differential equations that you need to grapple with... They're really freaking hard to solve.
>
> — *Grant Sanderson*

In this final results-focused chapter, I fulfil the promise set out in chapters 4 and 5, and derive closed-form expressions for the duration of suppression and dominance in tCFS, as well as for the degree of hysteresis separating the threshold for the suppression of a stimulus from awareness, and the breakthrough of a stimulus into awareness. Importantly, the resulting expression leads to a series of novel predictions, the first of which we test and confirm in a pre-existing psychophysical dataset.

This chapter is close to my heart; it is the first project I have ever worked on where a large portion of the work was done on pencil and paper (or more accurately, tablet and scribe). I spent months pretending to be a real mathematician, sitting in cafes scribbling notes and checking my (often incorrect) calculations with Mathematica. Most of this work turned out to be superfluous; all that the problem needed in the end was a little geometric intuition and a slightly exotic complex-valued function. It was a whole lot of fun.

6.1 Introduction

When incompatible stimuli are presented to each eye, binocular fusion is prevented and the stimulus presented to one eye is consciously experienced, whilst the other is suppressed from awareness [1–4]. Suppression is, however, transitory, and the two rival stimuli stochastically enter and disappear from awareness with a characteristic right-skewed distribution of dominance durations [5, 6]. Visual rivalry resulting from interocular conflict, therefore, offers

a psychophysically tractable window into the dynamical mechanisms underlying the transition between perceptual awareness and suppression, attracting considerable attention from experimentalists [7–21], and neural modellers [22–36] alike.

In visual rivalry paradigms that utilise stimuli with constant contrast, the minimal physiological conditions for awareness and suppression are theoretically well understood [27, 29, 37]. In binocular rivalry (the simplest and most common form of interoccular rivalry), two conflicting static stimuli are presented to each eye, and perception alternates between the two mutually exclusive stimulus percepts [2]. Here, adaptation and intrinsic noise both play an essential role [23, 34, 36, 38–41]. Models poised at the boundary between noisy excursions around a stable adaptation-driven limit cycle, and noise-driven jumps between adaptation-modulated fixpoint attractors provide the best fit to empirical data [39].

The relationship between the physical properties of the stimuli entering each eye and the duration of the resulting mutually exclusive rival percepts is remarkably consistent and can be captured by a small set of law-like propositions known as Levelts laws [42]. The propositions have proven extremely robust, receiving only minor modification in the 60 years since their conception [43]. Levelts laws have, therefore, served as an explanatory benchmark for computational models of binocular rivalry. As a result, the dynamical basis of the laws is, likewise, theoretically well understood. Closed-form expressions for each law can be derived from models with physiologically meaningful parameters [24, 29, 37], offering a quantitative explanation of the relative dominance of each percept in terms of the underlying dynamical processes and their coarse-grained physiological implementation.

The minimal conditions for awareness and suppression in continuous flash suppression (CFS; [44]), another canonical rivalry paradigm, can also be understood in terms of the same underlying principles. In CFS, a mask consisting of a sequence of high-contrast random patterns is presented to one eye and a static target stimulus is presented to the other. Unlike binocular rivalry where stochastic perceptual alternations occur every few seconds, CFS typically entails a single asymmetric period of dominance, whereby the dynamic random pattern of the mask is perceived for tens of seconds. Observers typically report the moment the weaker, previously suppressed visual target breaks free from suppression and enters awareness. Importantly, the prolonged period of mask dominance is thought to be enabled through a reduction in adaptation. Indeed, Shimaoka and Kaneko [45] showed that modifying a model of binocular rivalry so that the mask periodically drives different orientation selective populations leads to lengthened periods of suppression, on the order of tens of seconds, consistent with CFS.

The central appeal of visual rivalry paradigms is that although two visual stimuli enter the visual system, only one is consciously experienced whilst the other is suppressed, offering a practical way to experimentally study the mechanisms underlying visual awareness and suppression. However, in nearly all CFS studies, the threshold for awareness and suppression is confounded with the accumulation of adaptation. Although it is possible to study the dynamics of perceptual awareness and suppression, neither paradigm offers a reliable or practical means of measuring the threshold for awareness or suppression.

To overcome this limitation, Alais and colleagues [46] introduced a novel variant of CFS known as tracking continuous flash suppression (tCFS) which captures periods of suppression and awareness during continuous flash suppression. In tCFS, a high contrast dynamic mask is presented to one eye, whilst the contrast of the target stimulus presented to the other eye changes (log) linearly as a function of the observers perceptual report. The target stimulus starts at high contrast, dominating perception, and steadily decreases in contrast until the observer

reports that it has been suppressed from awareness by the mask. Upon report the contrast of the target stimulus reverses direction and increase until it breaks through suppression into awareness (continuing in a cycle). Stimulus contrasts registered at points of suppression and breakthrough provide threshold estimates for both breakthrough (i.e., awareness) and suppression. Moreover, the estimates are made quickly so that the effect of adaptation is reduced. In breakthrough, for example, instead of waiting until the target breaks suppression, the target rises in contrast until the breakthrough point is reached. The threshold reached in this way will be less affected by adaptation. Intriguingly, stimulus contrast thresholds measured at points of breakthrough and suppression show a robust pattern of hysteresis (known as suppression depth in the visual rivalry literature), with a higher contrast threshold required for a stimulus to breakthrough suppression into awareness, than to be suppressed from awareness. In addition, the depth of hysteresis was shown to increase as a function of contrast rate, pointing to a systematic role for adaptation in setting the threshold for awareness and suppression. tCFS thus offers an opportunity to test the generalisability of computational models of interocular competition

We develop a theoretical account of the neural mechanisms that govern awareness and suppression in tCFS. In chapter 5, we generalised a minimal model of binocular rivalry [29] to tCFS, and used numerical simulations to show that a singular mechanism, competitive inhibition between slowly adapting monocular populations, can account for perceptual awareness and suppression in both paradigms. Here, we use the same model to develop a quantitative account of the mechanisms governing awareness and suppression. Specifically, we extend the quantitative study of visual rivalry to the threshold for awareness and suppression by deriving closed-form expressions for dominance and suppression durations, and for the degree of hysteresis separating breakthrough and suppression thresholds, as a function of model parameters. We then leverage the expression for the depth of hysteresis to propose a novel empirical prediction, which we confirm in previously collected psychophysical data [46]. Finally, we derive two additional behavioural predictions that can be tested in future experiments.

6.2 Chapter 6 methods

The dynamics of the model are described by a system of four nonlinear ordinary differential equations (ODEs) with: reciprocal inhibition between competing monocular populations; self-adaptation; and percept-dependent stimulus contrast ramping (Figure 6.1A). This combination guarantees alternations between mutually-exclusive perceptual states,

$$\begin{aligned}
\tau_E \dot{E}_M &= -E_M + f(M + \varepsilon E_M - a E_S - g_M H_M) \\
\tau_H \dot{H}_M &= -H_M + E_M \\
\tau_E \dot{E}_S &= -E_S + f(S + \varepsilon E_S - a E_M - g_S H_S) \\
\tau_H \dot{H}_S &= -H_S + E_S \,,
\end{aligned} \tag{6.1}$$

where the four state variables E_M, E_S and H_M, H_S represent the aggregate excitatory neural activity of orientation-selective monocular populations driven by the mask and stimulus (denoted by the subscripts M and S respectively), and the aggregate slow hyperpolarising adaptation current of each population. For analytic tractability, the transfer function is a simple threshold nonlinearity $f(x) = \max(x, 0)$. The parameters ε, a, and g, are dimensionless and represent the strength of recurrent excitation, competitive inhibition, and adaptation, respectively. The population time constants τ_E and τ_H are in units of milliseconds. Each

neuronal population receives an independent (dimensionless) external drive (M, S). Unlike binocular rivalry, where the external drive shown to both eyes is stationary, in tCFS, a dynamic (flashing) mask is presented to one eye and the contrast of the target stimulus presented to the other eye changes (log) linearly with time. To approximate the effect of the dynamic mask whilst keeping the model analytically tractable, we drove the mask population with a constant stimulus but reduced the strength of adaptation in line with previous hypotheses [46] and modelling [45]. Model parameter values and numerical simulation details are supplied in Appendix 1 and (unless stated otherwise) are the same as those used by Wilson [29]. To model the percept-dependent cyclic linear increase and decrease in stimulus contrast, we represented the stimulus with a piecewise linear ODE with dynamics that depend on which population is dominant:

$$\dot{S} = \begin{cases} -\gamma & \text{if } E_S > E_M \\ \gamma & \text{if } E_S < E_M \,, \end{cases} \tag{6.2}$$

where γ is a dimensionless rate constant. We chose values of γ so that the difference between the maximum and minimum contrast rate values used by Alais and colleagues [46] was preserved. We matched γ to empirical data by converting the empirical contrast rates (originally in units of dB per frame with a refresh rate of 60Hz) to units of dB per millisecond and used the difference between the maximum and minimum contrast rates (0.0021 - 0.0063 dB/ms) to define the range over which γ varied. We then multiplied the full range of γ values by a constant on the interval $(10^0 - 10^{-4})$, and found the value (10^{-2}) that resulted in simulated dominance durations with the same order of magnitude as the empirical data.

Instead of analysing the dynamics of the full, four-dimensional nonlinear system, we took advantage of the separation of timescales between the (fast) neuronal and (slow) adaptation variables (i.e., $\tau_H \gg \tau_E$; [47, 48]). This allowed us to partition the dynamics into two phases, both of which could be studied on the phase plane. In the first instantaneous phase $H_L, H_R \approx 0$, and in the second asymptotic phase, $H_L \approx E_L$ and $H_R \approx E_R$. In the initial instantaneous phase of the dynamics, if $E_L, E_R > 0$, equation 6.1 can be reduced to a two-dimensional system with Jacobian matrix:

$$J_{E_M, E_S > 0} = \frac{1}{\tau_E} \begin{bmatrix} -1 + \varepsilon & -a \\ -a & -1 + \varepsilon \end{bmatrix} . \tag{6.3}$$

Geometrically, the existence of mutually exclusive perceptual states requires that the two asymptotically stable (mutually exclusive) states (corresponding to $E_M > 0, E_S = 0$ and $E_M = 0$ and $E_S > 0$) are separated by a saddle node. That is, the eigenvalues of $J_{E_M, E_S > 0}$ must have different signs. The eigenvalues are $\lambda_1 = \tau_E^{-1}(-1 - a + \varepsilon)$, and $\lambda_2 = \tau_E^{-1}(-1 + \varepsilon + a)$ for all physiologically plausible parameter values $a, \varepsilon > 0$ and $a > \varepsilon$, and as such $\lambda_1 < 0$, leading to the following inequality for $\lambda_2 > 0$:

$$a > 1 - \varepsilon \,. \tag{6.4}$$

If the inequality in equation 6.4 is satisfied, the model will converge to one of the two winner-take-all states. As adaptation dynamics never reduce neural activity below the rectification threshold of the transfer function, once the system is in a winner-take-all state, the state will be stable as long as the competing population is suppressed. Perceptual switches occur when the competing population escapes suppression [23, 38–41]. For example, if the mask-driven population E_M is dominant, and the stimulus-driven population E_S is suppressed, a perceptual switch will occur when E_S escapes suppression and inhibits E_M, which, for the

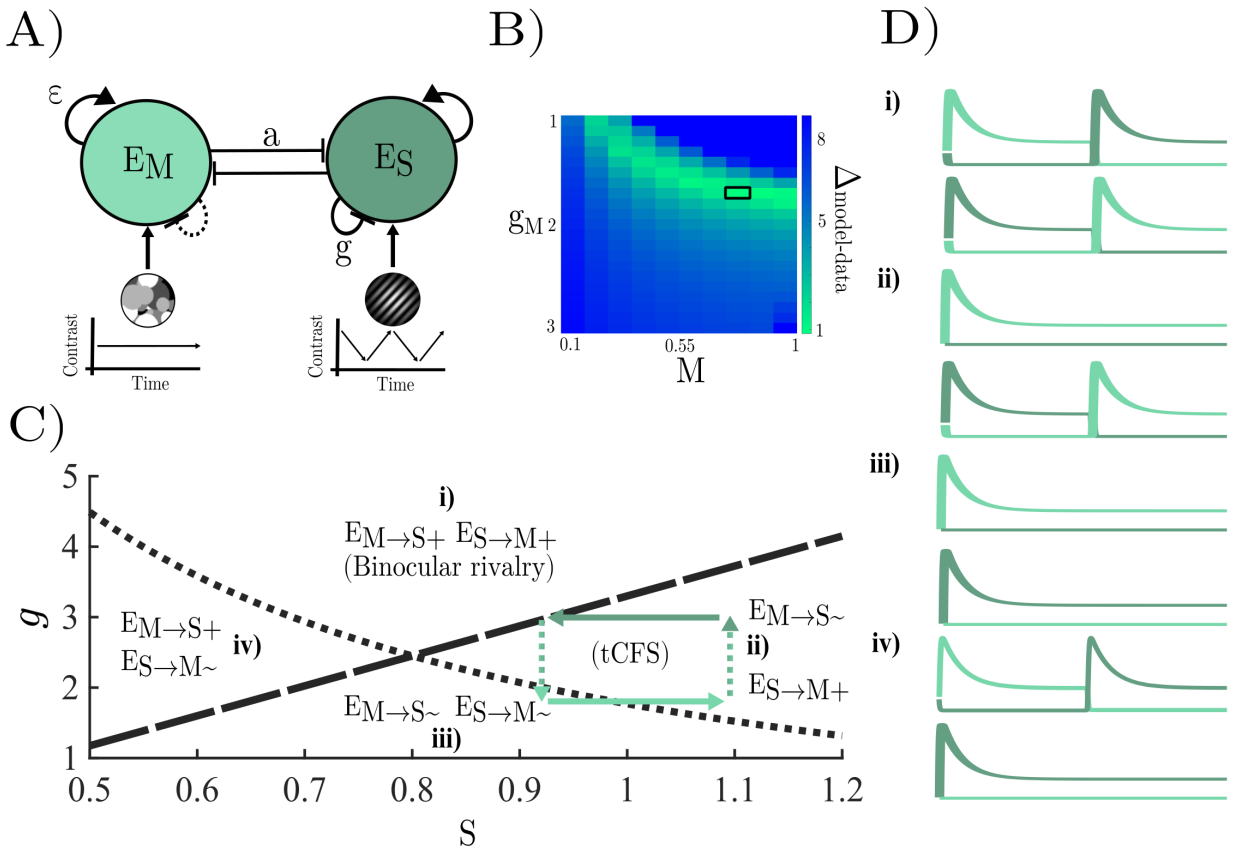

Figure 6.1: **A)** tCFS model architecture monocular populations driven by a mask (M) with constant contrast, and target stimulus (S) with percept-dependent ramping contrast, compete through mutual inhibition with asymptotic dynamics governed by adaptation. **B)** Dominance duration loss ($\Delta_{\text{model-data}}$) quantifying the fit between empirical and simulated dominance durations across the $[g_M \times M]$ parameter space. The best fitting parameter combination $g_M = 1.7, M = 0.8$ is highlighted by the solid black rectangle. **C)** The model $[g \times S]$ parameter space is partitioned into four distinct dynamical regimes by the inequalities in equation 6 (g_M- dotted line, g_S- dashed line): regime i) both populations escape suppression asymptotically via adaptation; regime ii) if the mask population is dominant, and the stimulus is suppressed, adaptation alone will not result in a perceptual switch, but if the stimulus is dominant the mask will asymptotically escape suppression via adaptation; regime iii) both the mask and target stimulus population cannot escape from suppression by adaptation alone; regime iv) if the stimulus population is dominant and the mask population is suppressed, adaptation alone will not result in a perceptual switch, but if the mask is dominant, the stimulus population will escape suppression via adaptation. An example trajectory across a full tCFS dominance-suppression cycle is shown in the green arrows. The dark green arrow represents periods where the stimulus-driven population is dominant. The stimulus starts at a high contrast value and gradually decreases in contrast until the mask-driven population (light green) escapes from suppression and becomes dominant, at which point the contrast of the stimulus starts to increase until the population driven by the stimulus escapes from suppression (and so on). **D)** Stylised representation of each dynamical regime.

simple threshold nonlinearity, occurs when the argument of the population transfer function $f(x)$ passes through zero from below. We refer to the argument of $f(x)$ as aggregate synaptic drive D:

$$\begin{aligned} D_M &= M + \varepsilon E_M(t) - aE_S(t) - g_M H_M(t), \\ D_S &= S + \varepsilon E_S(t) - aE_M(t) - g_S H_S(t)\,, \end{aligned} \tag{6.5}$$

where D takes negative values when the population is suppressed and positive values when dominant. There are two types of perceptual switches that can occur within the model:

adaptation-driven switches and stimulus-driven switches. Adaptation-driven switches occur when $g_{M/S}$ is strong enough to reduce the asymptotic firing rate of the dominant population to such a degree that the suppressed population is able to escape from inhibition. Such dynamics are likely solely governed by local circuit interactions. In contrast, input-driven switches occur when the self-inhibitory effect of adaptation is not sufficient for the suppressed population to escape from inhibition, and an increase in external drive to the suppressed population or a decrease in external drive (or possibly an increase/decrease in input from another cortical or subcortical region [16]) to the dominant population is necessary for a switch to occur. We can find the value of g separating these two regimes for each population by demanding that $D > 0$ and solving for the value of $g_{M/S}$ necessary for the suppressed population to escape from inhibition. For the asymptotic escape of the stimulus-driven population (E_S) from suppression by the mask-driven population (E_M), we obtain the following inequality:

$$g_M > a\frac{M}{S} - 1 + \varepsilon\,, \tag{6.6}$$

where we have relied on the fact that at a perceptual switch the model is in the second asymptotic phase of its dynamics, where $H_S \approx 0$, and $H_M \approx E_M$, and when E_S is suppressed $\varepsilon E_S = 0$. We find the equilibrium value of the (dominant) mask-driven population by substituting $H_M = E_M$ into the equation for $\dot{E}_M = 0$, and solve for E_M leading to $E_M(\infty) = M/(1 + g_M - \varepsilon)$, which when substituted into the expression for D_S in equation 6.5 gives the inequality in equation 6.6 assuming that adaptation in the suppressed population has sufficiently decayed (close to zero) before the switch occurs. The comparable expression for the asymptotic escape of the mask-driven population is $g_S > a\frac{S}{M} - 1 + \varepsilon$.

This leaves us with two inequalities defining the boundaries of four distinct dynamical regimes (i.e., switch type Œ monocular population) in the model's parameter space (Figure 6.1C). As suppression durations in CFS (∼10 50 s) are much greater than the timescale of the adaptation current (∼1 s), we hypothesised that the dynamics of the model that best represent tCFS would be located in the stimulus driven switching regime when the mask driven population is dominant (corresponding to regimes ii or iii in Figure 6.1C), and be in the adaptation-driven switching regime when the stimulus driven population is dominant (i.e., regimes i or ii in Figure 6.1C). As we did not want to simply assume this to be the case, we constrained adaptation and external drive values for the mask-driven population with behavioural data from Alais and colleagues [46].

Following Wilson [29], we fixed the adaptation current parameter for the stimulus-driven population by finding the value of g_S that resulted in an asymptotic firing rate that is approximately 30% of the maximum firing rate at stimulus onset, as is observed empirically [49]. We used this value as the upper bound on the strength of adaptation for the mask-driven population which we fit empirically. Specifically, we ran a grid search over the $g_M \times M$ parameter space to find the parameter value combination that best minimised the difference ($\Delta_{\text{model-data}}$) between simulated and empirical dominance durations across slow (0.0021 dB/ms), intermediate (0.0042 dB/ms), and fast (0.0063 dB/ms) contrast rates (Figure 6.1B). In the model, these stimulus values correspond to $\gamma = 0.0021, 0.0042, 0.0063 \times 10^{-2}$. The best fitting parameter combination was $g_M = 1.7, M = 0.8$. In line with our hypothesis, the trajectory of the best fitting parameters remained within regimes ii) and iii) (Figure 6.1C). All results reported below were stable across parameter combinations adjacent to the minima of $\Delta_{\text{model-data}}$.

6.3 Chapter 6 results

6.3.1 Hysteresis and contrast rate

In tracking continuous flash suppression (tCFS; [46]), the target stimulus is presented to one eye, and the flashing mask with constant contrast is presented to the other. The target starts at full contrast, and decreases log linearly (in units of dB) until the target stimulus is suppressed from awareness and the mask dominates perception. This perceptual reversal is indicated by a button press at which point the direction of contrast change reverses, continuing in a cycle until the trial is terminated (typical trials consist of 16 - 20 perceptual reversals; Figure 6.2A-B).

Two key empirical findings emerge from this paradigm. First, the contrast threshold for breakthrough and suppression is hysteretic. Across three experiments with multiple stimulus categories, the average difference between breakthrough and suppression thresholds was shown to be constant, with a reliably higher threshold for breakthrough than suppression. That is, the same stimulus, with the same contrast could dominant the content of awareness, or be suppressed from awareness, depending on the participants history of perceptual stimulation. Second, the degree of hysteresis is dependent on the rate of contrast change, with faster rates of contrast change leading to larger differences in breakthrough and suppression thresholds. This increasing difference in thresholds as a function of contrast rate points to a role for adaptation in accounting for the depth of hysteresis (Figure 2A-B and 2E; [46]).

To understand these results mechanistically, we simulated tCFS across 30 contrast rates evenly distributed across an interval proportional to the difference between the maximum and minimum contrast rates that were used empirically [46]. Each trial started by driving the model with a mask (with constant contrast) and an initially high contrast target stimulus. As was seen empirically, the stimulus-selective population won the initial competition for dominance, and remained dominant until the contrast of the target stimulus decreased sufficiently for the mask-driven population to escape suppression and inhibit the stimulus-selective population into silence (Figure 6.2C-D). At this point, the target stimulus started to increase in contrast again until the external drive entering the stimulus selective population was sufficient for the population to breakthrough suppression from the mask. Across all contrast rates, the model converged to an equilibrium with stationary breakthrough and suppression thresholds. Initial differences between successive threshold values, due to the target stimulus starting with high contrast, decayed away as adaptation approached its equilibrium value across successive cycles of breakthrough and suppression.

In line with empirical data, in our simulations, the depth of hysteresis in contrast thresholds increased as a linear function of contrast rate. With our analytical approach, we can explain this trend quantitatively by deriving expressions for the value of the target stimulus at breakthrough (S_B) and suppression (S_S) points. At the point of perceptual breakthrough, the dominant mask-driven population is well approximated by its asymptotic value $E_M(S_B) \approx E_M(\infty) = M/(1 + g_M - \varepsilon)$, and the stimulus-driven population is suppressed $E_S = 0$. Substituting these values into equation 6.5 and setting $D_S = 0$, we can then solve for S_B leading to:

$$S_B = a\frac{M}{1 + g_M - \varepsilon} + g_S H_S(T_{\text{Suppressed}})\,. \tag{6.7}$$

Similarly, at the point of perceptual suppression $E_S(T_{\text{Suppressed}}) \approx E_S(\infty) = S_S/(1 + g_S - \varepsilon)$,

and $E_M = 0$, which when substituted into $D_M = 0$, leads to:

$$S_S = \frac{1 + g_S - \varepsilon}{a}(M - g_M H_M(T_{\text{Dominant}}))\,. \tag{6.8}$$

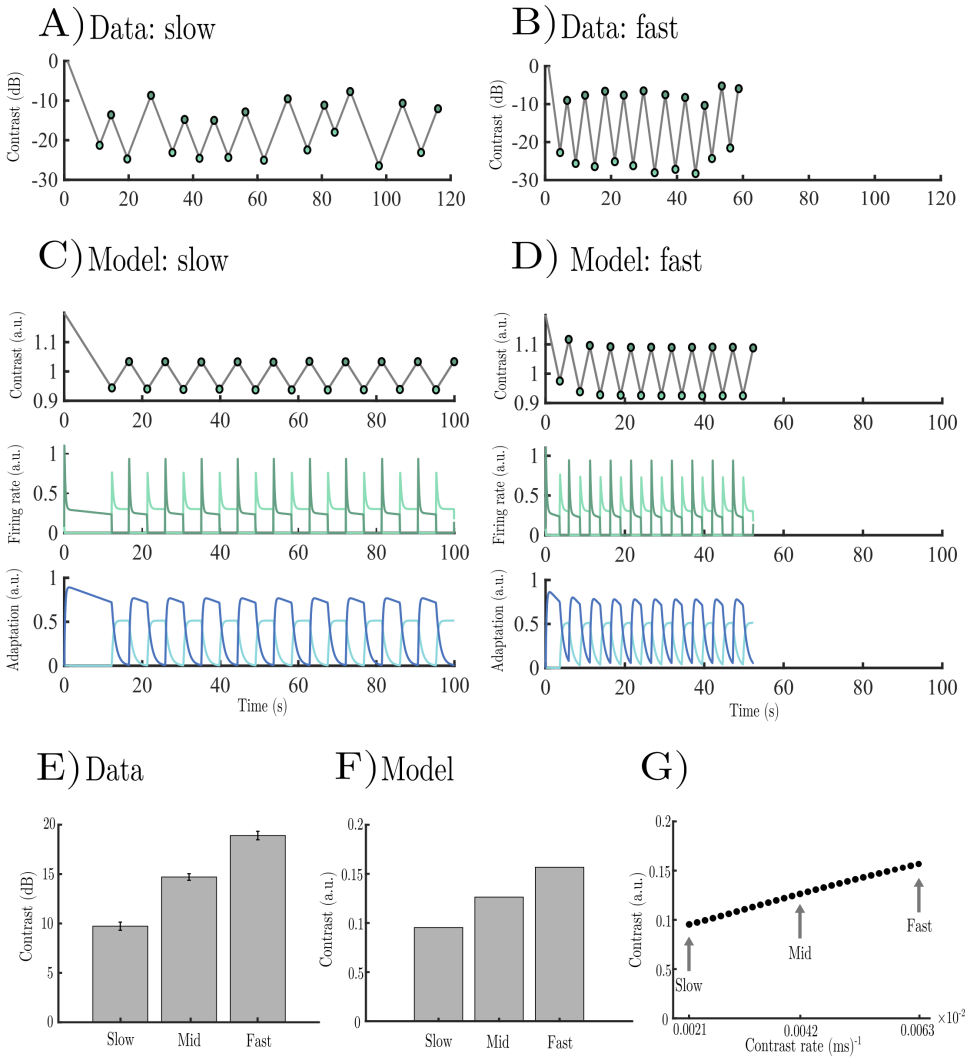

Figure 6.2: **A)** Stimulus contrast dynamics from two example tCFS trials with slow and fast contrast rates adapted from [46]. Points of suppression where the mask becomes dominant are highlighted with light green dots and points of perceptual breakthrough are highlighted with dark green dots. Each trial terminated after 20 reversals. **B-C)** Example simulated contrast dynamics, firing rates, and adaptation current dynamics for the population driven by the mask (light green) and the target stimulus (dark green). To aid in comparison with empirical data only 20 reversals are shown. We note that this is for purposes of illustration only, all simulated trials lasted for 120s. **D)** Approximate bifurcation diagrams with example model trajectories for the population driven by the target stimulus (upper dark green), and the population driven by the mask (lower light green). **E)** Empirical data for hysteresis depth across slow, mid, and fast contrast rates adapted from Alais and colleagues [46]. **F)** Simulated hysteresis depth for matched slow, mid, and fast contrast rates. **G)** Hysteresis depth across the full range of contrast rates used in simulation.

For slow contrast rates, dominance durations are relatively long and the exponential decay governing adaptation dynamics in the suppressed population approaches zero before the perceptual switch occurs (Figure 6.2C). Breakthrough and suppression thresholds are, therefore, dominated by the constant terms in equations 6.7-6.8. Indeed, for the lowest contrast rate

(0.0023×10^{-2} ms^{-1}), the constant terms in the expressions for breakthrough and suppression thresholds ($S_B - S_S \approx M(\frac{a}{1+g_M-\varepsilon} - \frac{1+g_S-\varepsilon}{a}) = 0.097$) provide a reasonable approximation to the degree of hysteresis observed in simulation ($S_B - S_S = 0.095$). In this regime, the degree of hysteresis is small and chiefly determined by the strength of asymptotic inhibition from the dominant population. As the contrast rate increases, dominance durations decrease, and the time-dependent adaptation terms begin to play a non-negligible role (Figure 6.2D), leading to a breakdown of the simple stationary approximation and an increase in hysteresis. To solve equations 6.7-6.8 for the general non-stationary case, we need an expression for dominance and suppression durations.

6.3.2 Dominance duration and the depth of hysteresis

Perceptual switches occur when the aggregate synaptic drive of the suppressed population pass through the rectification threshold from below and the population escapes suppression by overcoming competitive inhibition from the dominant population. We can, therefore, solve for dominance and suppression durations by substituting the exact solution for the dynamics of the stimulus (equation 6.2; $S(t) = S_B - \gamma t$ for $E_S > E_M$, and $S(t) = S_S + \gamma t$ for $E_S < E_M$) along with asymptotic approximations to the activity of the dominant population ($E_M(\infty), E_S(\infty)$) into the expression for aggregate synaptic drive (equation 6.5) yielding:

$$\begin{aligned} D_S &= S_S + \gamma t - a\frac{M}{1+g_M-\varepsilon} - g_S\frac{S_S}{1+g_S-\varepsilon}e^{-t/\tau_H} = 0\,, \\ D_M &= M - a\frac{S_B-\gamma t}{1+g_S-\varepsilon} - g_M\frac{M}{1+g_M-\varepsilon}e^{-t/\tau_H} = 0\,, \end{aligned} \tag{6.9}$$

where we have taken advantage of the fact that when a population is suppressed there is no recurrent activity ($\varepsilon E_{M/S} = 0$) and adaptation undergoes simple exponential decay with initial conditions given by the asymptotic activity of the now suppressed population (i.e. $H_{M/S}(t) = E_{M/S}(\infty)e^{-t/\tau_H}$). S_B and S_S are the values of the stimulus at points of breakthrough and suppression, which serve as initial conditions for the piecewise ODE governing the dynamics of the stimulus (equation 6.2). In essence, then, we can think of each successive period of dominance and suppression as setting the initial conditions for the following period.

Although we cannot solve for t in terms of elementary functions, both expressions can be shown to have exact solutions written in terms of the Lambert W (i.e. product log) function:

$$T_{\text{Suppressed}} = \underbrace{\frac{1}{\gamma}\left(\frac{aM}{1+g_M-\varepsilon} - S_S\right)}_{\text{Stationary drive}} + \underbrace{\tau_H W_0\left(\frac{g_S}{\gamma\tau_H}\frac{S_S}{1+g_S-\varepsilon}e^{-\frac{1}{\gamma\tau_H}\left(\frac{aM}{1+g_M-\varepsilon}-S_S\right)}\right)}_{\text{Time-dependent drive}}, \tag{6.10}$$

$$T_{\text{Dominant}} = \underbrace{\frac{1}{\gamma}\left(S_B - \frac{(1+g_S-\varepsilon)}{a}M\right)}_{\text{Stationary drive}} + \underbrace{\tau_H W_0\left(\frac{g_M M(1+g_S-\varepsilon)}{a\gamma\tau_H(1+g_M-\varepsilon)}e^{-\frac{1}{\gamma\tau_H}\left(S_B-\frac{(1+g_S-\varepsilon)}{a}M\right)}\right)}_{\text{Time-dependent drive}},$$

where W_0 is the principle branch of the Lambert W function. Both equations provide an excellent fit to dominance and suppression durations derived from numerical simulations across the full range of contrast rates (Figure 6.3A). Full derivation of equation 6.10 is supplied in Appendix 2

The first term in both expressions captures the stationary contribution of the difference between the excitatory drive entering the suppressed population and the level of asymptotic inhibition from the dominant population. In fact, if we neglect the time-dependent effects of adaptation, we can derive the stationary component of equation 6.10 by substituting the exact solution for the stimulus dynamics into equations 6.7-6.8 and solving for t. This term is a monotonically decreasing function of contrast rate proportional to γ^{-1}. Notably, in isolation, because the term does not capture any time-dependent effects, including those associated with adaptation, it systematically underestimates the durations of dominance and suppression.

The second term in both expressions captures the time-dependent effects of adaptation. To intuitively understand the role of contrast rate in governing the durations of dominance and suppression (equation 6.10), we need to take into account two key facts. First, for input > 0, the Lambert W function is monotonically increasing. Second, the argument of the Lambert

W function has the form $\frac{c_1}{\gamma}e^{-\frac{c_2}{\gamma}}$ where c_1 and c_2 are real-valued constants, which for all parameters used in the simulation ($\gamma << 1$), is also a monotonically increasing function of contrast rate. This term captures the fact that as contrast rates increase, there is less time for adaptation to recover, increasing the breakthrough threshold and decreasing the suppression threshold, meaning that it takes longer for input to travel between the breakthrough and suppression values. The closed-form expressions for dominance and suppression duration can, therefore, be seen as a solution to the trade-off between the opposing effects of contrast rate on dominance and suppression duration. As contrast rate increases, it takes less time for the stimulus to change from the breakthrough value to the suppression value (and vice versa), but by the same token, because adaptation has an additive effect on the breakthrough threshold, and a subtractive effect on the suppression threshold (see equations 6.7-6.8), as dominance durations decrease, the distance between breakthrough and suppression values increases.

Closed-form expressions for dominance and suppression duration in hand, we can now derive an expression for hysteresis depth as a function of contrast rate. Specifically, the depth of hysteresis is given by contrast rate and the length of the dominance/suppression period. This leads to the following two equations for the depth of hysteresis between successive periods of dominance and suppression:

$$\begin{aligned}\mathcal{H}_{S\to B} &= \gamma T_{\text{Suppressed}} \\ &= \underbrace{\left(\frac{aM}{1+g_M-\varepsilon} - S_S\right)}_{\text{Stationary drive}} + \underbrace{\gamma\tau_H W_0\left(\frac{g_S}{\gamma\tau_H}\frac{S_S}{1+g_S-\varepsilon}e^{-\frac{1}{\gamma\tau_H}\left(\frac{aM}{1+g_M-\varepsilon}-S_S\right)}\right)}_{\text{Time-dependent drive}} \\ \mathcal{H}_{B\to S} &= \gamma T_{\text{Dominant}} \\ &= \underbrace{\left(S_B - \frac{(1+g_S-\varepsilon)}{a}M\right)}_{\text{Stationary drive}} + \underbrace{\gamma\tau_H W_0\left(\frac{g_M M(1+g_S-\varepsilon)}{a\gamma\tau_H(1+g_M-\varepsilon)}e^{-\frac{1}{\gamma\tau_H}\left(S_B-\frac{(1+g_S-\varepsilon)}{a}M\right)}\right)}_{\text{Time-dependent drive}},\end{aligned} \tag{6.11}$$

where $\mathcal{H}_{S\to B}$ and $\mathcal{H}_{B\to S}$ denote the hysteretic difference between successive contrast thresholds between points of suppression and breakthrough and vice versa. Crucially, once the model has converged to an equilibrium, the expressions become interchangeable (i.e. $\mathcal{H}_{S\to B} = \mathcal{H}_{B\to S}$) and provide an excellent fit to the hysteresis depth values derived from numerical simulation (Figure 6.3B). One can understand the symmetry between periods of dominance and suppression by noting that:

$$\begin{aligned}S_{S_{n+1}} &= S_{B_n} - \gamma T_{\text{Dominant}_n}, \\ S_{B_{n+1}} &= S_{S_{n+1}} + \gamma T_{\text{Suppressed}_{n+1}},\end{aligned} \tag{6.12}$$

which, because of the successive dependence of breakthrough and suppression thresholds, leads to,

$$\begin{aligned}S_{S_{n+1}} &= S_{S_n} + \gamma T_{\text{Suppressed}_n} - \gamma T_{\text{Dominant}_n}, \\ S_{B_{n+1}} &= S_{B_n} - \gamma T_{\text{Dominant}_n} + \gamma T_{\text{Suppressed}_{n+1}},\end{aligned}$$

implying that once the model has converged to equilibrium, where $S_{B_{n+1}} = S_{B_n}$ and $S_{S_{n+1}} = S_{S_n}$, it must be the case that $T_{\text{Dominant}_n} = T_{\text{Suppressed}_n}$ (Figure 6.3A-B). We can understand this symmetry geometrically by treating each perceptual switch from dominance to suppression

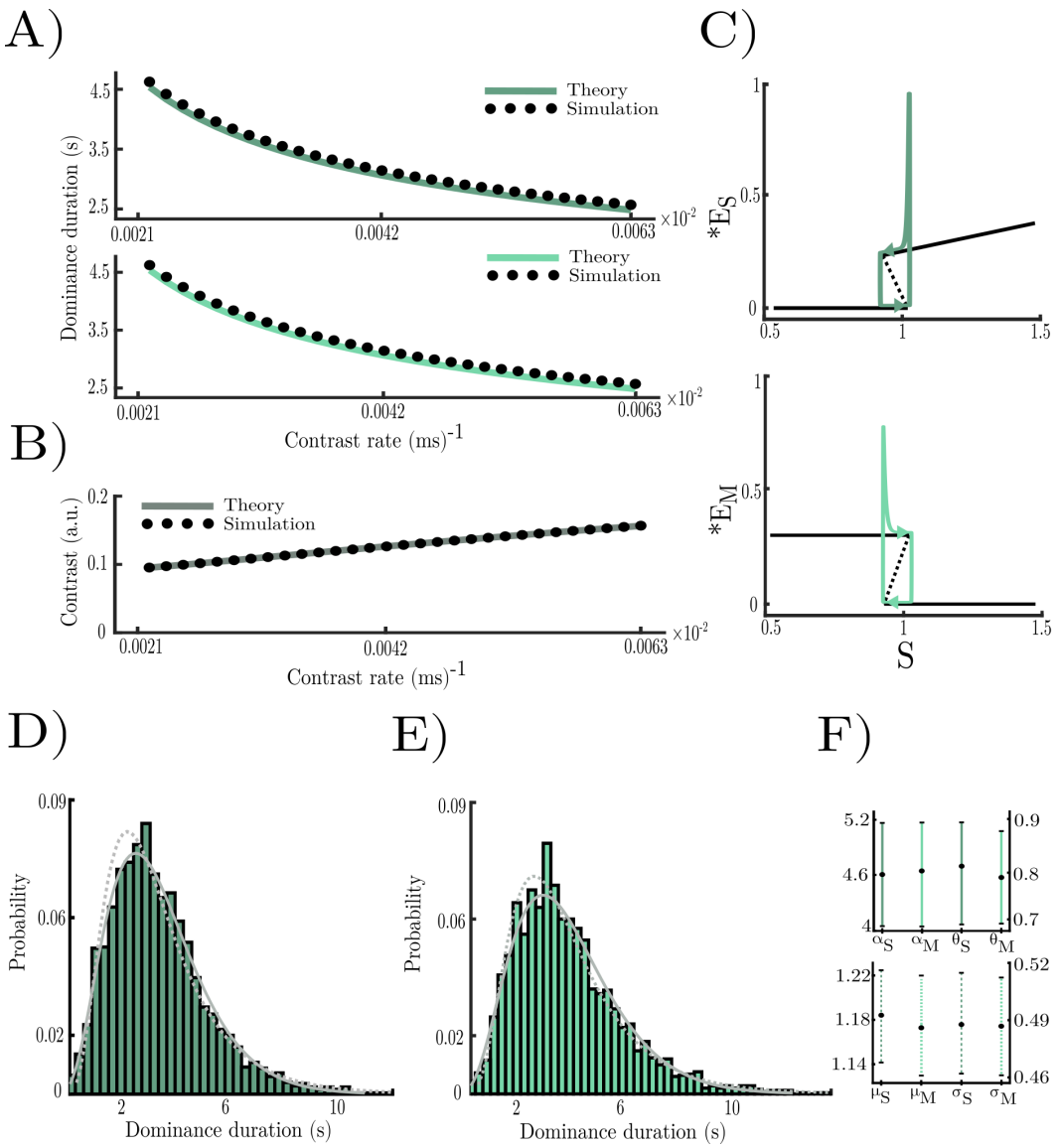

Figure 6.3: **A)** Simulated dominance (dark green) and suppression (light green) durations across the full range of contrast rates. Numerical simulation results are shown in black dots, and values predicted from analytic theory are shown in solid lines. **B)** Simulated hysteresis depth as a function of contrast rates. Numerical simulation results are shown in black dots, and values predicted from analytic theory are shown with solid lines. **C)** Approximate bifurcation diagrams with example model trajectories for the population driven by the target stimulus (upper dark green), and the population driven by the mask (lower light green). **D-E)** Distribution of dominance (dark green) and suppression (light green) durations aggregated across all three experiments from Alais and colleagues [46]. Grey lines show the fit of lognormal (dashed) and gamma (solid) distributions fit via maximum likelihood estimation. **F)** Parameters fit with 95% confidence intervals for dominance and suppression duration distributions for gamma (upper) and lognormal (lower) distributions. Left y-axes shows α and μ values right y-axes show θ and σ values.

(and vice versa) as a (saddle node) bifurcation and constructing bifurcation diagrams for each population with bifurcation points given by S_B and S_S (Figure 6.3C). Because the distance between S_B and S_S is constant across periods of dominance and suppression (once the model has reached equilibrium), the time it takes for the stimulus to decrease from S_B to S_S is identical to the time it takes for the stimulus to increase from S_S to S_B.

6.3.3 Confirmation of novel prediction

Crucially, the symmetry of dominance and suppression durations is a prediction that we can test in the empirical data of Alais and colleagues [46]. If we assume that the empirical data are generated by a noisy version of an equivalent dynamical system, the distribution of dominance durations should be approximately equal to the distribution of suppression durations. To test this prediction, we took a two-pronged approach. First, we aggregated the suppression and dominance duration data across all three experiments from Alais and colleagues [46]. We excluded trials with dominance or suppression durations longer than the time it takes for the stimulus to move from the minimum to the maximum allowed contrast value, as well as individual periods of dominance or suppression shorter than 500ms, as these instances likely reflect either a noise-driven perceptual switch or a failure to follow task instructions. In addition, as the prediction applies to systems close to their equilibrium value, we excluded the first two periods of dominance and suppression after trial onset. We then used a two-sample Kolmogorov-Smirnov test against the null hypothesis that the data come from the same underlying distribution. The test failed to reject the null hypothesis at the 5% significance level ($p = 0.5555$).

Second, having failed to reject the null hypothesis that the data come from the same underlying distribution, we next sought to find confirmatory evidence for the equivalence of the dominance and suppression distributions. We therefore fit gamma and lognormal distributions, which have both previously been shown to provide a good fit to bistable perception data [5, 6, 27, 41, 50, 51], to the dominance and suppression duration data via maximum likelihood estimation. Both distributions provided an excellent fit to the data (Figure 6.3D-F) with the gamma distribution performing marginally better than the lognormal distribution in terms of the negative log likelihoods (dominant $l_{\text{gamma}}/l_{\text{lognormal}} = 0.9945$, suppressed $l_{\text{gamma}}/l_{\text{lognormal}} = 0.9971$)[but, also see, 51]. In line with the prediction from the model, parameter estimates for dominance and suppression distributions were near identical for both the gamma ($\alpha_{\text{Dom}} = 4.64, \theta_{\text{Dom}} = 0.784$; $\alpha_{\text{Sup}} = 4.63, \theta_{\text{Sup}} = 0.78$) and lognormal ($\mu_{\text{Dom}} = 1.18, \sigma_{\text{Dom}} = 0.49$; $\mu_{\text{Sup}} = 1.17, \sigma_{\text{Sup}} = 0.48$) distributions with almost entirely overlapping 95% confidence intervals (Figure 6.3F).

6.3.4 Additional predictions

Having shown that the model accurately reproduces existing data and tested a novel prediction in existing data, we next sought to move beyond the current tCFS paradigm by generating two additional predictions that can be tested in future experiments. Inspired by previous bistable perception experiments that pharmacologically manipulate baseline levels of inhibition by giving participants a GABA agonist [52], we simulated the same tCFS paradigm whilst simultaneously sweeping the degree of competitive inhibition (a) between values of 3.3-3.5 (Figure 6.4A). We stayed within this narrow range of values as the model is highly sensitive to changes in competitive inhibition and values outside this range lead to empirically unrealistic behaviour. In line with the predictions of equation 6.10, increasing competitive inhibition increased the depth of hysteresis across the full range of contrast rates (Figure 6.4B), an effect that is chiefly mediated by increasing the stationary effect of inhibition. Averaged across $\mathcal{H}_{S\rightarrow B}$ and $\mathcal{H}_{B\rightarrow S}$, the relative contribution of each term to the total degree of hysteresis changed from stationary = 54.2%, time-dependent = 45.8% for $a = 3.3$ to, stationary = 89.61%, time-dependent = 10.39% for $a = 3.5$.

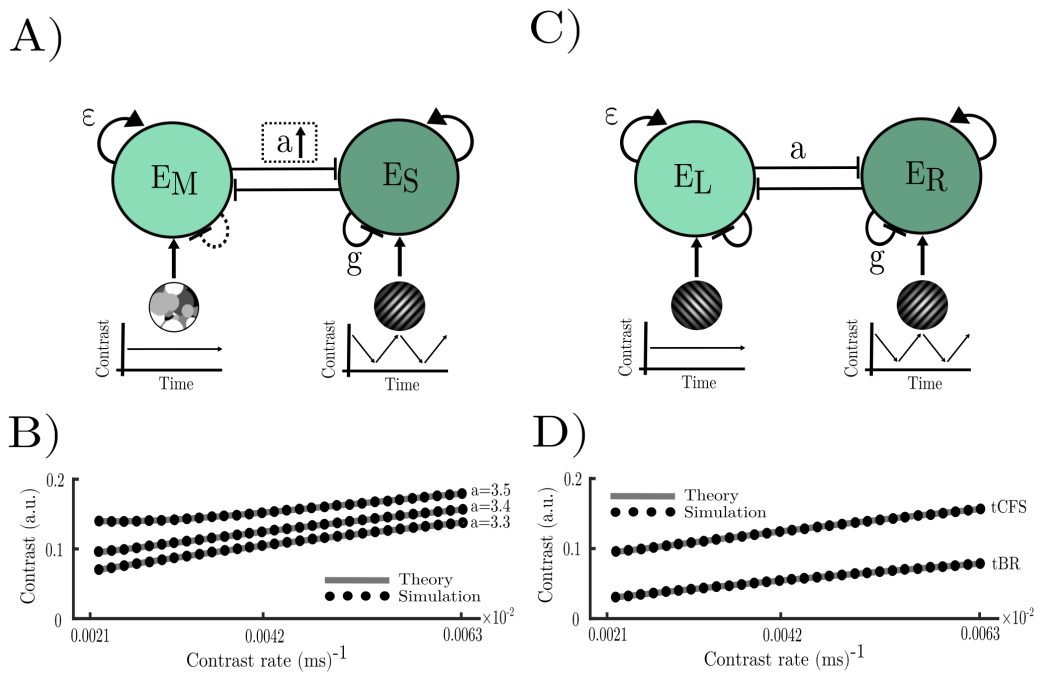

Figure 6.4: **A)** tCFS model architecture highlighting the role of competitive inhibition (a). **B)** Hysteresis depth for tCFS simulations across the full range of contrast rates with competitive inhibition values of 3.3, 3.4, and 3.5. **C)** Model architecture for tracking binocular rivalry (tBR). **D)** Hysteresis depth for tCFS and tBR simulations across the full range of contrast rates.

Next, we took advantage of the fact that manipulating the stimuli entering each eye provides an indirect means to manipulate the net strength of adaptation (Figure 6.4C). Specifically, instead of driving one eye with a flashing mask (as is done in standard tCFS), which periodically drives unadapted neurons, reducing the overall effect of adaptation [45], we drive both eyes with standard grating stimuli, leaving the contrast of one eye fixed and dynamically ramping the contrast of the other. This equalises the adaptation strength of each population. We refer to this extension of the tCFS paradigm as tracking binocular rivalry (tBR). We simulated tBR in the model by setting $g_M = g_S$. tBR almost halves the depth of hysteresis seen in standard tCFS (Figure 6.4D), an effect predicted by equation 6.10 that is primarily mediated by a reduction in the stationary effect of inhibition due to the increased accumulation of adaptation in the dominant population, reducing the asymptotic firing rate of the population driven by the fixed contrast stimulus. This reduces the breakthrough threshold when the dynamic stimulus-driven population is suppressed, which in turn reduces the stationary effect of inhibition when the same population is dominant, leading to a net decrease in the depth of hysteresis. Again, averaging across $\mathcal{H}_{S\to B}$ and $\mathcal{H}_{B\to S}$, the relative contribution of each term to the total degree of hysteresis changed from stationary = 76.8%, time-dependent = 23.2% for tCFS, to stationary = 39.22%, time-dependent = 60.78% for tBR.

6.4 Discussion

Building on chapter 5, which generalised Wilson's [29] minimal model of binocular rivalry to tracking continuous flash suppression (tCFS), we developed a quantitative account of the mechanisms governing awareness and suppression in tCFS. Visual rivalry is one of the few experimental paradigms in psychology and neuroscience where there are law-like explana-

tions and accompanying closed-form expressions, analogous to those observed in the physical sciences [see also, 53]. The key contribution of this paper is the derivation of closed-form expressions for the duration of dominance and suppression periods for the type of non-stationary stimuli used in tCFS, and an accompanying expression for the depth of hysteresis (i.e., suppression depth) between breakthrough and suppression thresholds.

In addition to explaining existing data, the expression for the depth of hysteresis led to the prediction that the distributions of dominance and suppression durations should be equivalent, which we tested and confirmed, in previously collected psychophysical data from Alais and colleagues [46]. Crucially, we also used the model and the accompanying expression for the depth of hysteresis to move beyond existing data and expose the model to potential falsification in two future experiments. Specifically, pharmacological manipulation of competitive inhibition should increase suppression depth, and equalising the strength of adaptation between competing neural populations by driving each population with standard grating stimuli instead of using a flashing mask should approximately halve the depth of hysteresis between breakthrough and suppression thresholds.

Lastly, the dynamical principles we employed in this study are similar to a wide range of multistability tasks. For instance, previous studies based on computational modelling [23, 54] suggested that the population-level neural dynamics underlying binocular rivalry [23] are similar to the dynamical regime underlying binocular flash suppression (BFS; [54]). In chapter 5, we demonstrated that this is the case for binocular rivalry and tCFS. And [45] showed that the same principles that apply to binocular rivalry apply to standard CFS. Given that all such circuit models use similar dynamical principles, parameters learned from any of these paradigms (BFS, BR, CFS, or tCFS) should also inform the neural dynamics and switching behavior of others. Thus, we believe, the neurodynamical implications of this theoretical study will likely go beyond tCFS, and give insight into a wide range of rivalry-like paradigms.

In sum, the model presented in this paper extends insights from quantitative models of binocular rivalry [29], and CFS [45], to tCFS. In tCFS, unlike binocular rivalry or standard CFS, it is possible to measure of the threshold for awareness and suppression, and depth of perceptual hysteresis separating the two thresholds, an effect we explain quantitatively in terms of the trade-off between the (relatively) stationary effects of inhibition and external drive, and the time-dependent effects of adaptation.

6.5 Chapter 6 supplementary material

6.5.1 tCFS model parameters

All simulations were run by integrating equation 6.1 numerically with the parameters in table 1 using the forward Euler method with a time step of $dt = 0.1$ ms in MATLAB 2023b.

Parameter	Description	Value	Units
τ_E	Neuronal population time constant	15	ms
τ_H	Hyperpolarizing adaptation current time constant	1000	ms
M	Constant drive of mask	0.8	a.u.
S	Initial condition for stimulus drive	$S_{t_0} = 1.2$	a.u.
ε	Strength of excitatory recurrent projections	0.05	a.u.
a	Strength of competitive inhibition	3.3-3.5 (3.4)	a.u.
g_M	Strength of mask population adaptation current	1.7	a.u.
g_S	Strength of stimulus population adaptation current	3	a.u.
γ	Rate of contrast change	$(0.0021 - 0.0063) \times 10^{-2}$	a.u.

Table 6.1: Parameter description, values, and units of the model described by equations 1-6 for tCFS simulations. Base parameter values used in parameter sweeps shown in parentheses

6.5.2 Derivation of dominance and suppression duration formula

Although it is not trivial to solve for t in equation 6.9, based on the assumptions and approximations described above, both equations can be shown to have the form $y = xe^x$ which for real x and y, on the restricted domain $x \geq 0$, have an inverse function W_0 (i.e. $x = W_0(xe^x)$), where W_0 is the principle branch of the Lambert W function.

Starting with line one of equation 6.9, where for compactness of notation we let $\bar{E}_M = E_M(\infty) = M/(1 + g_M - \varepsilon)$ and $\bar{E}_S = E_S(\infty) = S_S/(1 + g_S - \varepsilon)$, we have:

$$
\begin{aligned}
S_S + \gamma t - a\bar{E}_M - g_S\bar{E}_S e^{-t/\tau_H} &= 0 \\
a\bar{E}_M - S_S - \gamma t &= -g_S\bar{E}_S e^{-t/\tau_H} \\
(a\bar{E}_M - S_S - \gamma t)e^{t/\tau_H} &= -g_S\bar{E}_S \\
\left(t + \frac{S_S - a\bar{E}_M}{\gamma}\right) e^{t/\tau_H} &= \frac{g_S\bar{E}_S}{\gamma} \\
\left(\frac{t}{\tau_H} + \frac{S_S - a\bar{E}_M}{\gamma\tau_H}\right) e^{\frac{t}{\tau_H} + \frac{S_S - a\bar{E}_M}{\gamma\tau_H}} &= \frac{g_S\bar{E}_S}{\gamma\tau_H} e^{\frac{S_S - a\bar{E}_M}{\gamma\tau_H}} \\
\frac{t}{\tau_H} + \frac{S_S - a\bar{E}_M}{\gamma\tau_H} &= W_0\left(\frac{g_S\bar{E}_S}{\gamma\tau_H} e^{\frac{S_S - a\bar{E}_M}{\gamma\tau_H}}\right).
\end{aligned}
$$

After minor rearrangement and expanding both $\bar{E}_M$ and $\bar{E}_S$ we obtain,

$$T_{\text{Suppression}} = \underbrace{\left(\frac{aM}{1+g_M-\varepsilon} - S_S\right)}_{\text{Stationary drive}} + \underbrace{\gamma\tau_H W_0\left(\frac{g_S}{\gamma\tau_H}\frac{S_S}{1+g_S-\varepsilon}e^{-\frac{1}{\gamma\tau_H}\left(\frac{aM}{1+g_M-\varepsilon}-S_S\right)}\right)}_{\text{Time-dependent drive}}.$$

The derivation of the dominance time from line two of equation 6.9 is slightly more complex, as changes in the strength of the stimulus have a delayed effect on the subthreshold dynamics of the mask-driven population. This is because the strength of asymptotic inhibition from the dominant population is dependent on adaptation. To compensate for the delay, we added a correction term to the expression for asymptotic firing rate $\tilde{S}_B = S_B - \gamma\tau_{\text{Delay}}$. We derived the value of the correction term by finding the value of τ_{Delay} which best minimised the difference between the closed-form expression for the asymptotic firing rate at suppression points ($E_S(\infty) = (\tilde{S}_B - \gamma t)/(1+g_S-\varepsilon)$) and the value derived from simulation. We averaged this value across contrast rates, giving $\tau_{\text{Delay}} = 760$ ms for the tCFS simulations, and $\tau_{\text{Delay}} = 634$ ms for the tBR simulations.

Correction in hand we arrive at,

$$\begin{aligned}
M - \frac{a(\tilde{S}_B-\gamma t)}{1+g_S-\varepsilon} - g_M\bar{E}_M e^{-t/\tau_H} &= 0 \\
M - \frac{a\tilde{S}_B}{1+g_S-\varepsilon} + \frac{a\gamma t}{1+g_S-\varepsilon} &= g_M\bar{E}_M e^{-t/\tau_H} \\
\left(M - \frac{a\tilde{S}_B}{1+g_S-\varepsilon} + \frac{a\gamma t}{1+g_S-\varepsilon}\right)e^{t/\tau_H} &= g_M\bar{E}_M \\
\left(\frac{t}{\tau_H} - \frac{a\tilde{S}_B - M(1+g_S-\varepsilon)}{a\gamma\tau_H}\right)e^{\frac{t}{\tau_H}-\frac{a\tilde{S}_B-M(1+g_S-\varepsilon)}{a\gamma\tau_H}} &= \frac{g_M\bar{E}_M(1+g_S-\varepsilon)}{a\gamma\tau_H}e^{-\frac{a\tilde{S}_B-M(1+g_S-\varepsilon)}{a\gamma\tau_H}} \\
\frac{t}{\tau_H} - \frac{a\tilde{S}_B - M(1+g_S-\varepsilon)}{a\gamma\tau_H} &= W_0\left(\frac{g_M\bar{E}_M(1+g_S-\varepsilon)}{a\gamma\tau_H}e^{-\frac{a\tilde{S}_B-M(1+g_S-\varepsilon)}{a\gamma\tau_H}}\right),
\end{aligned}$$

expanding $\bar{E}_M$ and solving for t, we obtain,

$$T_{\text{Dominance}} = \underbrace{\frac{1}{\gamma}\left(\tilde{S} - \frac{(1+g_S-\varepsilon)}{a}M\right)}_{\text{Stationary drive}} + \underbrace{\tau_H W_0\left(\frac{g_M M(1+g_S-\varepsilon)}{a\gamma\tau_H(1+g_M-\varepsilon)}e^{-\frac{1}{\gamma\tau_H}\left(\tilde{S}-\frac{(1+g_S-\varepsilon)}{a}M\right)}\right)}_{\text{Time-dependent drive}}.$$

Each expression is dependent on the stimulus values at suppression (S_S) and breakthrough points (S_B), respectively. Instead of simply pulling these values from the simulation, we leveraged the successive dependence of breakthrough and suppression values, and the duration of suppression and dominance, to write equation 6.10 and equation 6.12 as a set of difference equations which we iterated until convergence (see, Algorithm 1).

Algorithm 1 Iterative update of S_B, $T_{\text{Dominance}}$, and $T_{\text{Suppression}}$

Initialise: $S_B \leftarrow S_{t_0}$

Initialise: $T_{\text{Dominance}} \leftarrow \frac{1}{\gamma}\left(S_B - \frac{(1+g_S-\varepsilon)}{a}M\right)$

repeat

$S_S \leftarrow S_B - \gamma T_{\text{Dominance}}$

$T_{\text{Suppression}} \leftarrow \frac{1}{\gamma}\left(\frac{aM}{1+g_M-\varepsilon} - S_S\right) + \tau_H W_0\left(\frac{g_S}{\gamma\tau_H}\frac{S_S}{1+g_S-\varepsilon}e^{-\frac{1}{\gamma\tau_H}\left(\frac{aM}{1+g_M-\varepsilon}-S_S\right)}\right)$

$S_B \leftarrow S_S + \gamma T_{\text{Suppression}}$

$T_{\text{Dominance}} \leftarrow \frac{1}{\gamma}\left(\tilde{S}_B - \frac{(1+g_S-\varepsilon)}{a}M\right) + \tau_H W_0\left(\frac{g_M M(1+g_S-\varepsilon)}{a\gamma\tau_H(1+g_M-\varepsilon)}e^{-\frac{1}{\gamma\tau_H}\left(\tilde{S}_B-\frac{(1+g_S-\varepsilon)}{a}M\right)}\right)$

until convergence.

We initialised the breakthrough value of the stimulus with the initial condition of the stimulus used in the simulations, and the initial dominance duration value with the stationary component of equation 6.10. The difference equations rapidly converged across all contrast rates to the equilibrium values shown in Figure 6.3 and Figure 6.4. This approach allowed us to arrive at closed-form expressions that predicted the behaviour of the model without the use of any information derived from the simulation results themselves (with the exception of the delay correction).

Chapter 6 references

[1] Randolph Blake. "A neural theory of binocular rivalry". In: *Psychological Review* 96.1 (1989), pp. 145–167. DOI: 10.1037/0033-295X.96.1.145.

[2] Randolph Blake. "A Primer on Binocular Rivalry, Including Current Controversies". In: *Brain and Mind* 2.1 (2001), pp. 5–38. DOI: 10.1023/A:1017925416289.

[3] Randolph Blake and Nikos K. Logothetis. "Visual competition". In: *Nature Reviews Neuroscience* 3.1 (2002), pp. 13–21. DOI: 10.1038/nrn701.

[4] Randolph Blake. "The perceptual magic of binocular rivalry". In: *Current Directions in Psychological Science* 31.2 (2022), pp. 139–146.

[5] R. Fox and J. Herrmann. "Stochastic properties of binocular rivalry alternations". In: *Perception & Psychophysics* 2.9 (1967), pp. 432–436. DOI: 10.3758/BF03208783.

[6] W. J. M. Levelt. "Note on the Distribution of Dominance Times in Binocular Rivalry". In: *British Journal of Psychology* 58.1-2 (1967), pp. 143–145. DOI: 10.1111/j.2044-8295.1967.tb01068.x.

[7] N. K. Logothetis, D. A. Leopold, and D. L. Sheinberg. "What is rivalling during binocular rivalry?" In: *Nature* 380.6575 (1996), pp. 621–624. DOI: 10.1038/380621a0.

[8] Jude F Mitchell, Gene R Stoner, and John H Reynolds. "Object-based attention determines dominance in binocular rivalry". In: *Nature* 429.6990 (2004), pp. 410–413.

[9] David Alais and David Melcher. "Strength and coherence of binocular rivalry depends on shared stimulus complexity". In: *Vision Research* 47.2 (2007), pp. 269–279. DOI: 10.1016/j.visres.2006.09.003.

[10] Randolph Blake and J. Camisa. "On the inhibitory nature of binocular rivalry suppression". In: *Journal of Experimental Psychology: Human Perception and Performance* 5.2 (1979), pp. 315–323. DOI: 10.1037/0096-1523.5.2.315.

[11] M. Hollins and G. Bailey. "Rivalry target luminance does not affect suppression depth". In: *Perception & Psychophysics* 30 (1981), pp. 201–203. DOI: 10.3758/BF03204480.

[12] C. Lunghi and D. Alais. "Congruent tactile stimulation reduces the strength of visual suppression during binocular rivalry". In: *Scientific Reports* 5.1 (2015), p. 9413. DOI: 10.1038/srep09413.

[13] V. A. Nguyen, A. W. Freeman, and D. Alais. "Increasing depth of binocular rivalry suppression along two visual pathways". In: *Vision Research* 43.19 (2003), pp. 2003–2008. DOI: 10.1016/S0042-6989(03)00314-6.

[14] O. Teng Leng and M. S. Loop. "Visual suppression and its effect upon color and luminance sensitivity". In: *Vision Research* 34.22 (1994), pp. 2997–3003.

[15] N. Tsuchiya et al. "Depth of interocular suppression associated with continuous flash suppression, flash suppression, and binocular rivalry". In: *Journal of Vision* 6.10 (2006), p. 6. DOI: 10.1167/6.10.6.

[16] Hagar Gelbard-Sagiv et al. "Human single neuron activity precedes emergence of conscious perception". In: *Nature communications* 9.1 (2018), p. 2057.

[17] Jan K Hesse and Doris Y Tsao. "A new no-report paradigm reveals that face cells encode both consciously perceived and suppressed stimuli". In: *eLife* 9 (2020), e58360.

[18] V. Kapoor et al. "Decoding internally generated transitions of conscious contents in the prefrontal cortex without subjective reports". In: *Nature Communications* 13 (2022), p. 1535.

[19] Alice Drew et al. "Conflict monitoring and attentional adjustment during binocular rivalry". In: *European Journal of Neuroscience* 55.1 (2022), pp. 138–153.

[20] Abhilash Dwarakanath et al. “Bistability of prefrontal states gates access to consciousness”. In: *Neuron* 111.10 (2023), pp. 1666–1683.

[21] Daniel P Montgomery et al. “Responses to conflicting binocular stimuli in mouse primary visual cortex”. In: *bioRxiv* (2025), pp. 2024–12.

[22] Peter Dayan. “A Hierarchical Model of Binocular Rivalry”. In: *Neural Computation* 10.5 (1998), pp. 1119–1135. DOI: 10.1162/089976698300017377.

[23] Panagiota Theodoni et al. “Cortical microcircuit dynamics mediating binocular rivalry: the role of adaptation in inhibition”. In: *Frontiers in human neuroscience* 5 (2011), p. 145.

[24] Carlo R. Laing and Carson C. Chow. “A Spiking Neuron Model for Binocular Rivalry”. In: *Journal of Computational Neuroscience* 12.1 (2002), pp. 39–53. DOI: 10.1023/A:1014942129705.

[25] R. Moreno-Bote, J. Rinzel, and N. Rubin. “Noise-Induced Alternations in an Attractor Network Model of Perceptual Bistability”. In: *Journal of Neurophysiology* 98.3 (2007), pp. 1125–1139.

[26] S. Safavi and P. Dayan. “A decision-theoretic model of multistability: Perceptual switches as internal actions”. In: *bioRxiv* (2024). DOI: 10.1101/2024.12.06.627286.

[27] A. Shpiro et al. “Dynamical Characteristics Common to Neuronal Competition Models”. In: *Journal of Neurophysiology* 97.1 (2007), pp. 462–473. DOI: 10.1152/jn.00604.2006.

[28] Hugh R. Wilson. “Computational evidence for a rivalry hierarchy in vision”. In: *Proceedings of the National Academy of Sciences* 100.24 (2003), pp. 14499–14503. DOI: 10.1073/pnas.2333622100.

[29] Hugh R. Wilson. “Minimal physiological conditions for binocular rivalry and rivalry memory”. In: *Vision Research* 47.21 (2007), pp. 2741–2750. DOI: 10.1016/j.visres.2007.07.007.

[30] S. J. Gershman, E. Vul, and J. B. Tenenbaum. “Multistability and Perceptual Inference”. In: *Neural Computation* 24.1 (2012), pp. 1–24. DOI: 10.1162/NECO_a_00226.

[31] Hugh R. Wilson. “Binocular contrast, stereopsis, and rivalry: Toward a dynamical synthesis”. In: *Vision Research* 140 (2017), pp. 89–95. DOI: 10.1016/j.visres.2017.07.016.

[32] H.-H. Li et al. “Attention model of binocular rivalry”. In: *Proceedings of the National Academy of Sciences* 114.30 (2017). DOI: 10.1073/pnas.1620475114.

[33] Pantelis Leptourgos et al. “A functional theory of bistable perception based on dynamical circular inference”. In: *PLOS Computational Biology* 16.12 (2020), e1008480.

[34] Robin Cao et al. “Binocular rivalry reveals an out-of-equilibrium neural dynamics suited for decision-making”. In: *elife* 10 (2021), e61581.

[35] Kenneth Barkdoll, Yuhua Lu, and Victor J Barranca. “New insights into binocular rivalry from the reconstruction of evolving percepts using model network dynamics”. In: *Frontiers in Computational Neuroscience* 17 (2023), p. 1137015.

[36] Yuxuan Wu, Liufang Xu, and Jin Wang. “Nonequilibrium dynamics and thermodynamics provide the underlying physical mechanism of the perceptual rivalry”. In: *Physical Review Research* 7.2 (2025), p. 023059.

[37] J. Seely and C. C. Chow. “Role of mutual inhibition in binocular rivalry”. In: *Journal of Neurophysiology* 106.5 (2011), pp. 2136–2150. DOI: 10.1152/jn.00228.2011.

[38] Jan W. Brascamp et al. “The time course of binocular rivalry reveals a fundamental role of noise”. In: *Journal of Vision* 6.11 (2006), pp. 8–8. DOI: 10.1167/6.11.8.

[39] A. Shpiro et al. "Balance between noise and adaptation in competition models of perceptual bistability". In: *Journal of Computational Neuroscience* 27.1 (2009), pp. 37–54. DOI: 10.1007/s10827-008-0125-3.

[40] Jochen Braun and Maurizio Mattia. "Attractors and noise: twin drivers of decisions and multistability". In: *Neuroimage* 52.3 (2010), pp. 740–751.

[41] Alexander Pastukhov et al. "Multi-stable perception balances stability and sensitivity". In: *Frontiers in computational neuroscience* 7 (2013), p. 17.

[42] W.J.M. Levelt. "On binocular rivalry". PhD thesis. Soesterberg, The Netherlands: Institute for Perception RVO-TNO, 1965.

[43] Jan W. Brascamp, P. Christiaan Klink, and Willem J. M. Levelt. "The 'laws' of binocular rivalry: 50 years of Levelt's propositions". In: *Vision Research* 109 (2015), pp. 20–37. DOI: 10.1016/j.visres.2015.02.019.

[44] N. Tsuchiya and C. Koch. "Continuous flash suppression reduces negative afterimages". In: *Nature Neuroscience* 8.8 (2005), pp. 1096–1101. DOI: 10.1038/nn1500.

[45] D. Shimaoka and K. Kaneko. "Dynamical systems modeling of Continuous Flash Suppression". In: *Vision Research* 51.6 (2011), pp. 521–528. DOI: 10.1016/j.visres.2011.01.009.

[46] David Alais et al. "tCFS: A new CFS tracking paradigm reveals uniform suppression depth regardless of target complexity or salience". In: *eLife* 12 (2024), e91019.

[47] Wulfram Gerstner et al. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. 1st. Cambridge University Press, 2014. DOI: 10.1017/CBO9781107447615.

[48] Xiao-Jing Wang. *Theoretical Neuroscience: Understanding Cognition*. CRC Press, 2025.

[49] D.A. McCormick. "Cholinergic and Noradrenergic Modulation of Thalamocortical Processing". In: *Trends in Neurosciences* 12 (1989), pp. 215–221.

[50] S. R. Lehky. "Binocular rivalry is not chaotic". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 259.1354 (1997), pp. 71–76.

[51] Jan W Brascamp et al. "Distributions of alternation rates in various forms of bistable perception". In: *Journal of vision* 5.4 (2005), pp. 1–1.

[52] A. M. van Loon et al. "GABA Shapes the Dynamics of Bistable Perception". In: *Current Biology* 23.9 (2013), pp. 823–827.

[53] Zhaoping Li. *Understanding vision: theory, models, and data*. Oxford University Press, 2014.

[54] Theofanis I Panagiotaropoulos et al. "A common neurodynamical mechanism could mediate externally induced and intrinsically generated transitions in visual awareness". In: *PLoS One* 8.1 (2013), e53833.

Chapter 7

Thesis Conclusion

We shall not cease from exploration
and the end of all our exploring
will be to come back
to the place from which we came
and know it for the first time

— *T.S. Eliot*

7.1 Where we have been and where are we going?

The aim of this thesis has been to build a quantitative bridge between the cellular-level mechanisms of perception discovered and characterised in animal models - transgenic mouse models in particular - and the behavioural signatures of perceptual awareness studied in human psychophysics. Whether implicit or explicit, this goal is shared by most practitioners in the neuroscience of consciousness, at least among those with a bent for quantitative theory. By this I mean simply that most of the major theories of consciousness appeal, in some way, to findings from both animal models and human participants (e.g. [1–3]). What sets the approach taken in this thesis apart is that I have focused exclusively on explaining, reproducing, and extending the results of a small set of concrete experimental paradigms using biophysical models jointly constrained by neurobiological data and psychophysical principles.

This stands in contrast to what I see as the current dominant approach in theorising about consciousness: starting with an existing theoretically-motivated characterisation of consciousness, such as global broadcasting in a frontoparietal network [4, 5], or finding the partition of a system that maximises integrated information [6, 7], and then building tractable models and metrics that explain the experimental results in terms of these existing theoretical constructs.

I suspect that most of the current prominent theories of consciousness started with an approach

analogous to the one I am arguing for. This is certainly the case for Global Workspace Theory. I first encountered the T.S. Eliot poem that opens this chapter as an undergraduate while reading Baars' 1988 book "A Cognitive Theory of Consciousness" [8], in which he introduced Global Workspace Theory. Baars took a piecemeal approach very similar to the one I am advocating for. Each chapter opened with a list of experimental findings he wanted to explain, and over the course of the book, he gradually built up to what would become Global Workspace Theory. Although there were no equations, he used concepts from symbolic artificial intelligence to give some formal context to his conceptual theory. This book left a lasting impression on me, and it is in homage to Baars' contribution to the field that I conclude this thesis with the same poem.

A lot has changed, however, in the decades since most of the prominent theories of consciousness were first proposed. The tools of modern systems neuroscience give us an unprecedented ability to record from [9, 10], and causally control [11, 12], neural activity. Theoretical neuroscience is also no longer the exclusive dominion of former physicists. Neuroscience now trains its own theorists [13–16], making mathematically sophisticated and neurobiologically informed modelling much more commonplace.

Enough has changed, therefore, both in terms of data and theoretical tools, that I think it is more useful to study individual experimental paradigms in their own right, rather than trying to fit new results into existing theories that were not designed with these results in mind. Indeed, if we buy the argument that the way forward for neuroscience is to construct cross-level mechanistic theories [15], constituted by quantitative models constrained by neurobiological and behavioural considerations across scales and levels of analysis, it would be truly shocking if existing theoretical constructs mapped neatly onto scales and experiments they were never designed to explain (cf. [17, 18]). To be clear, I am not advocating for the abandonment of existing theories of consciousness as they have been, and continue to be, useful. In science, as in life, it is almost always good to let a thousand blossoms bloom. Rather, I think we have a lot to gain by making cross-scale mechanistic modelling - agnostic to existing theories of consciousness - the norm rather than the exception. Once we have reliable quantitative models that explain data across scales and experiments, we will be in a place where it is useful to once again begin the process of constructing unifying explanatory theories.

Part 1 of this thesis started by reviewing cellular and systems-level contributions of thalamocortical loops to conscious state, and argued that the same matrix thalamus-$L5_{ET}$ circuit motif that controls conscious state gates the contents of perceptual awareness. We underscored vital, yet distinct, roles for core- and matrix-type thalamic neurons, and concluded that thalamocortical loops play a central role in controlling the global state and local contents of consciousness.

Having reviewed the state of the art in the systems and cellular neuroscience of consciousness and argued for a thalamocortical basis for perceptual awareness, in chapter 3, we put computational meat on the conceptual bones of chapter 2 by developing a biophysical model of perceptual awareness that incorporated the essential elements of the relevant mesoscale thalamocortical anatomy and cellular physiology. The model reproduced the key neural and behavioural signatures of perceptual awareness in a mouse model of threshold detection, including the animals' responses to causal perturbations of the thalamocortical circuit. We then generalised the model to visual rivalry, an experimental paradigm typically used to study perceptual awareness in human and non-human primates. Analysis of the model based upon dynamical systems theory revealed that the same thalamocortical mechanisms that governed

perceptual awareness in threshold detection - thalamus-gated bursting in $L5_{ET}$ neurons - determined perceptual dominance in visual rivalry. Crucially, in addition to reproducing existing neural and behavioural findings, the model generated testable electrophysiological predictions.

In part 2, we moved from a multi-compartment thalamocortical spiking model, with neuronal elements derived from cellular and systems neuroscience, to a more abstract and mathematically tractable cortical neural mass model. The construction of the spiking model in chapter 3 and its generalisation from threshold detection to visual rivalry relied upon the existence of psychophysical constraints, in the form of Levelt's laws [19, 20]. Showing that our model was consistent with these laws gave us sufficient confidence in the model's behaviour to proceed with prediction generation. Therefore, in part 2 of the thesis, I sought to derive closed-form law-like expressions for a recently proposed visual rivalry paradigm - tracking continuous flash suppression (tCFS; [21]) - that allows researchers to reliably estimate the contrast threshold for entry into, and suppression from, awareness. By deriving such constraints, it becomes possible to build spiking models of tCFS that may, in time, give a general cellular-level explanation of the threshold for awareness that can be tested in animal models.

Chapter 4 bridges the connection between the thalamocortical spiking model explored in part 1 (chapters 2-3) and the more abstract cortical neural mass model introduced in part 2 (chapters 5-6), I showed that we can derive - via a separation of time scale argument - the purely cortical neural mass model of rivalry from a more detailed thalamocortical neural mass model that obeys many of the same principles learned from the spiking model (e.g. matrix-thalamus controlled cortical gain). I showed that the purely cortical neural mass, derived from the full thalamocortical model, has near identical dynamics to the more detailed model, and that both models obey the closed-form expression for dominance duration derived from the cortical model. Although we did not derive the thalamocortical neural mass directly from the spiking model, this heuristic approach should lend some credence to the idea that principles learnt at this more abstract level will generalise to the spiking level.

In chapters 5 and 6, we explored the dynamical principles underlying the suppression of perceptual content from awareness, and the breakthrough of perceptual contents into awareness, in both binocular rivalry and tCFS. In chapter 5, we extended an existing dynamical model of binocular rivalry to encompass a threshold detection variant of binocular rivalry, as well as tCFS. Through numerical simulation, we showed that a single mechanism, competitive (hysteretic) inhibition between slowly adapting monocular populations, is sufficient to account for the suppression depth (i.e., degree of hysteresis in contrast thresholds) findings across both paradigms.

Having established the face validity of the model, in chapter 6, we derive closed-form expressions for the quantities qualitatively characterised in chapter 5, and move beyond existing experimental data to generate novel empirical predictions based on the model. Specifically, we quantitatively explained the hysteretic transition between periods of awareness and suppression in tCFS by deriving closed-form expressions for the duration of perceptual dominance and suppression, and for the degree of hysteresis (i.e., the depth of perceptual suppression), as a function of model parameters. Finally, we used the model to generate a series of novel behavioural predictions, the first of which - distributions of suppression and dominance durations in tCFS should be approximately equivalent - we tested and confirmed in human psychophysical data. In addition, although we do not discuss this in the chapter, I recently learned that the third prediction - the depth of hysteresis between breakthrough and suppression thresholds should approximately halve if identical stimuli are presented to either eye -

has subsequently also been confirmed [22]. tCFS, like binocular rivalry, therefore has dynamics that conform to a small set of law-like expressions, making it an excellent candidate for psychophysically constrained biophysically detailed modelling.

To summarise what we have learned: At the spiking level, perceptual thresholds are controlled by matrix-type projections to the apical dendrites and the apical-coupling zone, which controls the gain of $L5_{ET}$ cells. In threshold detection paradigms, this determines whether a stimulus can evoke sufficient activity to stand out against noisy background activity and be reliably read out by downstream processes. In visual rivalry, this same process determines whether a population can generate and maintain a sufficient level of recurrent excitation to suppress competing populations into silence via mutual inhibition. Assuming that the thalamus and apical dendrites are operating in a normal regime, thalamic conductances quickly approach their asymptotic values that, given the timescales involved, can be well approximated by a stationary gain parameter. At the level of neural masses (assuming gain rapidly approaches its equilibrium value), the threshold for content to break through into awareness, or be suppressed from awareness, in visual rivalry is chiefly determined by the relatively stationary effect of inhibition from the dominant population, and the time-dependent subtractive effect of adaptation. This cross-level mechanistic perspective blurs the distinction between enabling background conditions and causally relevant component processes; what appears like a background condition at one scale may be a causally relevant component process in the mechanism under study at another scale.

Across both the spiking and neural mass levels, the transition from unconscious to conscious processes is governed by saddle node bifurcations (i.e., when a stable and unstable fixed point collide and annihilate). At the spiking level, the transition from regular spiking to intrinsic bursting is governed by a saddle node bifurcation in the apical dendrites, with external drive from the somatic compartment and synapses acting as a bifurcation parameter. At the level of neural masses, the transition between suppression and awareness in visual rivalry is, likewise, governed by a saddle node bifurcation, but with the input from each eye (with strength determined by the contrast of the stimulus) acting as a bifurcation parameter. The key quantitative prediction of the models put forward in this thesis is, therefore, that we should see signatures of criticality (e.g., increasing variance as a function of distance to bifurcation) that scales according to a saddle node bifurcation, at the level of both single neurons and populations of neurons, as an organism transitions from being perceptually unaware to aware (assuming the bifurcation is approached sufficiently slowly).

In what remains of this chapter, I briefly address two issues that were put aside over the course of the thesis: when are neurobiological details important for modelling? And what is the relationship between the findings of the thesis and existing theories of consciousness? Finally, I close by speculating about what a more mature cross-level mechanistic thalamocortical theory of the properties of conscious contents might look like.

7.2 When do neurobiological details matter?

One may wonder, if reduced purely cortical models behave near identically to more detailed thalamocortical models that incorporate cellular detail, why bother studying the microscale implementation details? The answer is: we do not want to only understand the brain at the scale that is most amenable to compact explanation and prediction. We also want to be able to control the brain's dynamics. All psychiatric drugs have cellular mechanisms of action [23];

if we are ever going to understand how, why, and when such drugs work, we need to have models that include cellular-level mechanisms. Even if such mechanisms can be treated as parameters rather than variables in many contexts, we cannot simply assume that this is the case when we experimentally manipulate them. Perturbing the system through experimental intervention will, in all likelihood, move these variables far away from their equilibrium values.

The answer to the question - "*when do neurobiological details matter?*" - will, therefore, depend on two key factors: the first is the timescale of the neurobiological process under study, and the second is the spatial and temporal scale of the experimental intervention. When experimental interventions are isolated to macroscale variables, with stationary or relatively slow stimulus manipulations, it is safe to assume the relevant neurobiological processes will rapidly reach their equilibrium values and can, therefore, be safely treated as stationary parameters. In this setting, experimental psychologists can continue to work at a high level of abstraction with concepts such as gain control and competitive inhibition without worrying too much about how such processes are implemented. If, however, we want to experimentally manipulate cellular processes or understand how neuronal processes are implemented in the brain such that they can be causally manipulated, details of the implementation must be included in the model. Simply put, it is not possible to simulate causal interventions into model components that do not exist. If we ever want to control [24] the brain's dynamics, cellular and neurobiological detail cannot be ignored.

7.3 How does the work align with existing theories of consciousness?

As has been noted elsewhere [25, 26], the cellular- and circuit-level conditions for awareness explored in part 1 of this thesis fit well with many of the major neuronal theories of consciousness. The diffuse projections of the matrix-type thalamus may be a circuit-level mechanism underlying the non-linear and widespread "ignition" response proposed by Global Neuronal Workspace theory to underlie the transition from unconscious to conscious processing [5, 27–29]. The improvement in signal-to-noise ratio associated with bursting aligns with signal detection theoretic versions of Higher-order Theory [30], and the recurrent nature of matrix thalamus-$L5_{ET}$ loops could be considered a thalamocortical extension of the currently corticocentric Recurrent Processing Theory [31]. In addition, in previous work, our group has shown that diffuse matrix-type control of bursting on a sheet of $L5_{ET}$ cells maximises an approximate measure of integrated information [32], in line with Integrated Information Theory [6]. I suspect, therefore, that exploring the interaction between the cellular conditions for awareness interrogated in this thesis and the topology of large-scale brain networks will be of crucial importance in resolving the ongoing debate about the macroscale network conditions necessary for awareness.

7.4 What might a mature thalamocortical theory of conscious contents look like?

The primary focus of this thesis has been on how perceptual content becomes conscious, irrespective of the subjective properties of the specific content. In particular, my coauthors and I argued that the interactions between $L5_{ET}$ cells and matrix thalamic cells offer a plausible

mechanism for explaining how content becomes conscious. It is important to note, however, that we are not claiming that $L5_{ET}$ cells are the sole cortical basis of the content of consciousness. For instance, pyramidal neurons in supragranular layers with narrow selectivity profiles [33] are known to modulate the gain of $L5_{ET}$ cells [34] and, in so doing, are well placed to influence the nature of the content that becomes conscious. Core neurons in the thalamus, which make precise, driver-like connections to L2-4 intra-telencephalic neurons, are anatomically situated to fulfil a similar functional role. Few studies have directly investigated the potential role of core neurons in consciousness. Nonetheless, a mature cross-level thalamocortical theory of conscious contents should incorporate an appreciation of the nuanced neuroanatomy of the core (in addition to matrix) thalamic nuclei, and their embedding within the rest of the brain. Indeed, this may explain several key aspects of the contents of consciousness, including its *integrated* (i.e., unified) [35–38], *segregated* (i.e., differentiated) [39–41] and seemingly *continuous* nature [42–44] (Figure 7.1).

Conscious experiences in the typical waking state are *integrated*, meaning percepts are experienced as unified wholes [36, 40–42, 45], not disparate, disconnected component pieces (for exceptions, see [36, 45]). When perceiving a visual scene, we do not sequentially perceive individual features of objects; the contents of the scene are experienced *en masse* in a common visual space. With others [40–42], we (at least partially) attribute this property to the anatomical and functional properties of thalamocortical loops. In particular, the thalamus contains orders of magnitude fewer cells than the cerebral cortex [46, 47], meaning that information embedded within cortical circuits will, by necessity, be compressed when processed through the thalamus before returning to the cerebral cortex [23, 48, 49] (Figure 7.1A). We propose that this pervasive motif of information compression contributes to the unified, low-dimensional "gist" quality of conscious contents. Specifically, the low-dimensional nature of the representation could not resemble the rich, detailed content ever-present in visual experience; it might, however, give rise to the "gist" of scenes or events (e.g., informationally-compressed versions of the scene), perhaps contributing to less vivid, but still adaptive aspects of experience. Consistent with the neural aspects of this hypothesis, thalamic activity in humans has been shown to be temporally coincident with low-dimensional network integration in complex cognitive tasks [49–52]. The resolution of these data is not sufficient to resolve whether this occurred via core or matrix nuclei (or both); however, there is evidence that the expression of genetic markers differentiating core and matrix populations covaries with similar low-dimensional cortical patterns [53].

Another key feature of consciousness is that perceptual contents are typically *segregated* from one another and experienced as distinct, differentiated perceptual objects [39–41]. Successive conscious states are not exactly alike and are often mutually exclusive, suggesting the need for relatively isolated circuits that eliminate competing sources of information processing not currently involved in contributing to the contents of consciousness. What role might the thalamocortical loops play in the segregated nature of conscious experience (Figure 7.1B)? There are three key anatomical principles that support this capacity: i) there are almost no excitatory connections between thalamic projection neurons [54] (only the posterior intralaminar nuclei excite other excitatory thalamic neurons); ii) the inhibitory reticular nucleus is recruited in an activity-dependent fashion; and iii) individual cells within the reticular nucleus project in a "fan-out" fashion that ultimately inhibits more cells than were required to engage the inhibition. The confluence of these anatomical features results in a relatively global process of divisive normalization [23, 55–57]. If a thalamic neuron is coerced into spiking (whether by subcortical or cortical inputs), then as it projects back to the cerebral cortex, the excitatory

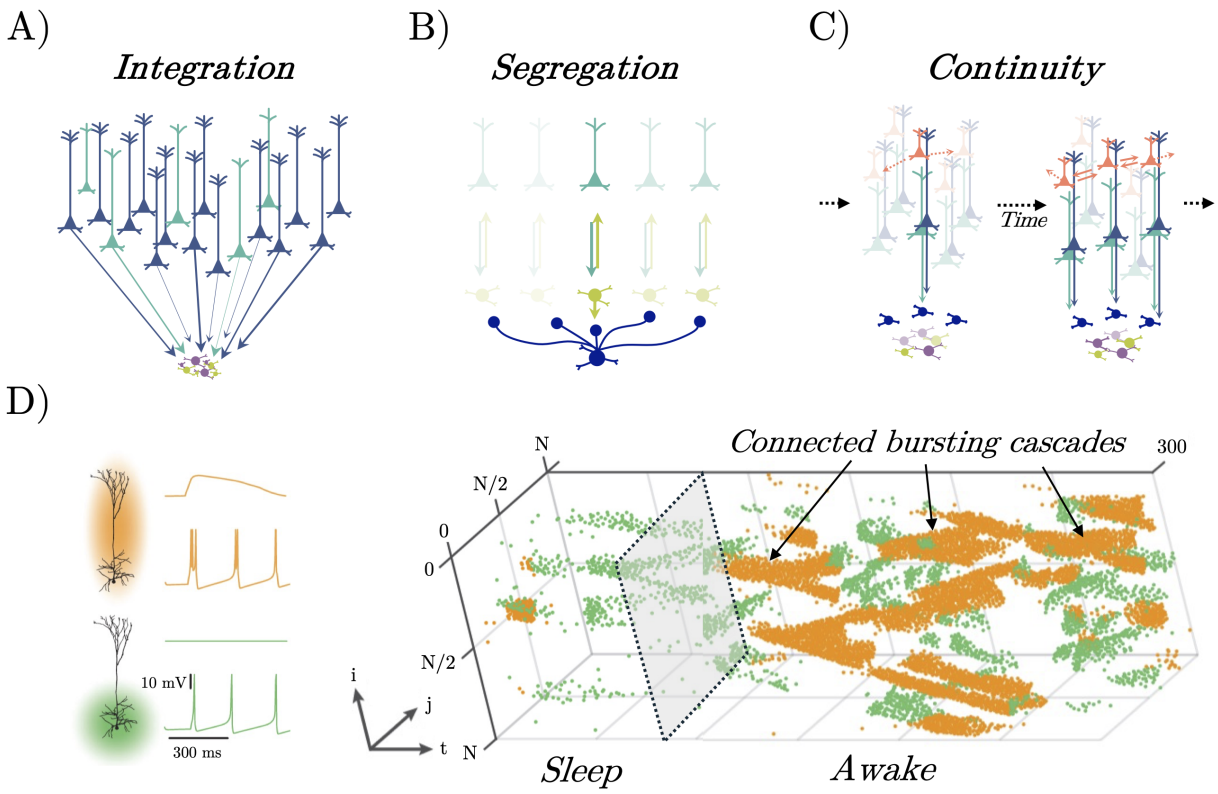

Figure 7.1: **Thalamic contributions to the character of conscious contents.** (**A-C**) The constraints placed on ongoing dynamics by the organization of the thalamus may help to explain key features of conscious information processing, such as: integration (**A**), which we argue arises as a product of the dimensionality reduction imposed by the converging inputs to the smaller number of cells in the thalamus (lower row), relative to the cerebral cortex (upper row); segregation (**B**), which is a natural byproduct of the divisive normalization - like process that arises from activity - dependent recruitment of the inhibitory thalamic reticular nucleus (dark blue); and continuity (**C**), wherein the combination of driver and modulatory roles in the thalamus imposes a Matthew effect' on cellular populations that provides an activity boost to neurons that are connected to the currently dominant thalamocortical ensemble. (**D**) A spiking neural model was created to mimic a key feature of $L5_{ET}$ cells namely, that they shift from a regular-spiking (green; input to basal dendrites only) to a burst-firing (orange; simultaneous activation of apical and basal dendrites) mode when gated by matrix thalamic inputs. After fitting the model to electrophysiological data from anesthetized/awake humans and naturally-sleeping macaques, the authors showed that the location in model parameter space that best fit to sleep was associated with predominantly regular spiking activity (green dots), whereas in the awake regime (following the gray dotted square), model $L5_{ET}$ neurons (embedded on a $N = 100 \times 100$ cortical sheet) were capable of entering into a burst-firing mode that, in turn, triggered burst-firing in synaptically connected neurons leading to sustained activity that could propagate coherently across the cortical sheet, a process labeled as "connected bursting cascades". Panel D adapted from [32] with permission. Note that these mechanisms are simply intended to convey plausible means (and not exhaustive) for enacting each feature.

cell also engages a relatively broad inhibitory process that ultimately hyperpolarizes any thalamic cells in the nearby anatomical vicinity. These cells thus have to obtain more excitatory drive in order to fire, making it less likely that they will form a part of the ongoing active coalition of neurons reverberating through thalamocortical loops that support the contents of consciousness.

A third crucial feature of typical conscious experience is the fact that transitions between conscious contents have a seeming *continuity* [42–44, 58] i.e., neural processes that can maintain coherence over protracted periods (Figure 7.1C). The most direct evidence for a core-type thalamic contribution to the experience of continuity comes from work on spatial constancy

the process underlying our stable perception of the world despite the saccadic eye movements that cause visual image displacement ∼3-4 times per second on average in primates. Spatial constancy is thought to arise from the corollary discharges associated with eye movements (i.e., copies of eye movement commands) that traverse from the superior colliculus to core-rich (calbindin-absent) lateral MD through to the cortical frontal eye fields [59]. This circuit enables remapping of the visual receptive field of frontal eye field neurons immediately before an eye movement from the current receptive field location to the future location (i.e., the new location of the receptive field after the movement). Such spatial remapping across eye movements is commonly thought to contribute to our visual perception of a stable world [60, 61]. Notably, prospective information about the upcoming saccade emerges earlier in the MD than in the frontal cortex [62], with peak activity immediately before movement initiation [63, 64]. This activity is movement-direction specific and transmitted to the frontal eye field for spatial remapping. Accordingly, muscimol-induced deactivation of these MD neurons prevented the remapping of frontal eye field neurons' receptive fields [61]. Moreover, MD deactivation perturbs corollary discharge processing and movement-related perception, but not the eye movement itself [65]. These experiments suggest that MD plays a crucial computational role in the stabilization of conscious contents.

Further (albeit weaker) evidence comes from the fact that MD has also been shown to play a vital role in working memory processes [66–69]. Although consciousness and working memory can be dissociated [70–73], the contents of working memory encoded in delay period activity are typically a part of the contents of consciousness [72]. In mice and primates, ongoing activity in the MD-frontal cortex loop has been shown to support task performance across delay periods; and deactivation of MD in mice and a thalamocortical computational model has been shown to prevent sustained spiking activity in frontal cortex and impair performance [74–77]. Whilst these studies were not designed to dissociate conscious contents from post-perceptual processes, they suggest core-cortex loops may underlie temporally-protracted, seemingly continuous aspects of conscious contents. Precisely-controlled experiments that manipulate specific cellular populations in nuanced perceptual contexts will allow more direct tests of this hypothesis.

Finally we speculate that by balancing integration and segregation thalamocortical coalitions evolve over time in a way that maintains the coherent character of typical experience. Specifically, recent modelling work from our group has shown that the extended thalamocortical circuit, when active, allows information to propagate across the cortical sheet through matrix-dependent bursting in coalitions of $L5_{ET}$ cells [32]. We speculate that the activity of the dominant coalition - reflecting the current content of consciousness - propagates across the cortical sheet through matrix primed cortico-cortical connections, whilst maintaining itself through reentrant interactions with core nuclei that also act to inhibit competing coalitions through the recruitment of the thalamic reticular nucleus (Figure 7.1D). This suggests that the dominant coalition may be able to leverage the "Matthew effect" of accumulated advantage [78]: that is, the thalamocortical coalition that is most active at a given time can enact a disproportionate influence [79] over the processes that define the *next* coalition that will be the most active [23], both because it can send spikes to all neurons connected to it, and because it does not have to compete with other potential challenger coalitions (all of which have been hyperpolarized due to the influence of the activity-dependent reticular nucleus; Figure 7.1B).

7.5 Concluding remarks

Irrespective of whether any of the particular computational models proposed in this thesis turn out to be correct, as animal models and the tools of systems neuroscience more generally are increasingly brought into close contact with the psychophysical methods used to study consciousness in humans, I think that the cross-level mechanistic [15] approach to neural modelling I have advocated for will become increasingly important. In particular, if we can quantitatively bridge the explanatory gap between the cellular mechanisms and psychophysical signatures of consciousness across a wide range of paradigms and model organisms, we may well be in a place where we can begin to construct a unified theory of consciousness that is useful and illuminating to philosophers, scientists, and clinicians alike.

Chapter 7 references

[1] INTREPID CONSORTIUM. *OSF Registries | Accelerating research on consciousness: An adversarial collaboration to test contrasting predictions of the Integrated Information Theory and Predictive Processing accounts of consciousness*. 2021. URL: https://osf.io/35rhx?mode=&revisionId=&view_only=.

[2] D. Pal et al. "Differential Role of Prefrontal and Parietal Cortices in Controlling Level of Consciousness". In: *Current Biology* 28.13 (2018), 2145–2152.e5. DOI: 10.1016/j.cub.2018.05.025.

[3] B. van Vugt et al. "The threshold for conscious report: Signal loss and response bias in visual and frontal cortex". In: *Science* 360.6388 (2018), pp. 537–542. DOI: 10.1126/science.aar7186.

[4] S. Dehaene. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Penguin Publishing Group, 2014.

[5] G.A. Mashour et al. "Conscious Processing and the Global Neuronal Workspace Hypothesis". In: *Neuron* 105 (2020), pp. 776–798.

[6] L. Albantakis et al. "Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms". In: *PLOS Computational Biology* 19.10 (2023), e1011465. DOI: 10.1371/journal.pcbi.1011465.

[7] G. Tononi et al. "Integrated information theory: From consciousness to its physical substrate". In: *Nature Reviews Neuroscience* 17.7 (2016), pp. 450–461. DOI: 10.1038/nrn.2016.44.

[8] B. J. Baars. *A Cognitive Theory of Consciousness*. Cambridge University Press, 1988.

[9] C. Grienberger et al. "Two-photon calcium imaging of neuronal activity". In: *Nature Reviews Methods Primers* 2.1 (2022), p. 67. DOI: 10.1038/s43586-022-00147-1.

[10] N. A. Steinmetz et al. "Challenges and opportunities for large-scale electrophysiology with Neuropixels probes". In: *Current Opinion in Neurobiology* 50 (2018), pp. 92–100. DOI: 10.1016/j.conb.2018.01.009.

[11] L. Fenno, O. Yizhar, and K. Deisseroth. "The Development and Application of Optogenetics". In: *Annual Review of Neuroscience* 34 (2011), pp. 389–412. DOI: 10.1146/annurev-neuro-061010-113817.

[12] D. J. Urban and B. L. Roth. "DREADDs (Designer Receptors Exclusively Activated by Designer Drugs): Chemogenetic Tools with Therapeutic Utility". In: *Annual Review of Pharmacology and Toxicology* 55 (2015), pp. 399–417. DOI: 10.1146/annurev-pharmtox-010814-124803.

[13] P. Dayan and L. F. Abbott. *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Massachusetts Institute of Technology Press, 2001.

[14] Eugene M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, 2006. DOI: 10.7551/mitpress/2526.001.0001.

[15] Xiao-Jing Wang. *Theoretical Neuroscience: Understanding Cognition*. CRC Press, 2025.

[16] H. R. Wilson. *Spikes, decisions, and actions: The dynamical foundations of neuroscience*. Oxford University Press, 1999.

[17] P. M. Churchland. "Eliminative Materialism and the Propositional Attitudes". In: *The Journal of Philosophy* 78.2 (1981), p. 67. DOI: 10.2307/2025900.

[18] P. M. Churchland. *Matter and Consciousness, third edition*. MIT Press, 2013.

[19] Jan W Brascamp, P Christiaan Klink, and Willem JM Levelt. "The laws of binocular rivalry: 50 years of Levelts propositions". In: *Vision research* 109 (2015), pp. 20–37.

[20] Willem JM Levelt. *On binocular rivalry*. Tech. rep. Institute for Perception, 1965.

[21] David Alais et al. "tCFS: A new CFS tracking paradigm reveals uniform suppression depth regardless of target complexity or salience". In: *eLife* 12 (2024), e91019.

[22] D. Alais et al. "A new 'tracking' version of Continuous Flash Suppression to quantify suppression strength: Constant CFS suppression for all image types & two times the suppression strength of binocular rivalry". In: *Journal of Vision* 25.9 (2025), p. 2886. DOI: 10.1167/jov.25.9.2886.

[23] J.M. Shine. "The Thalamus Integrates the Macrosystems of the Brain to Facilitate Complex, Adaptive Brain Network Dynamics". In: *Progress in Neurobiology* 199 (2021), p. 101951.

[24] E. J. Müller, B. R. Munn, and J. M. Shine. "The brain that controls itself". In: *Current Opinion in Behavioral Sciences* 63 (2025), p. 101499. DOI: 10.1016/j.cobeha.2025.101499.

[25] J. Aru, M. Suzuki, and M.E. Larkum. "Cellular Mechanisms of Conscious Processing". In: *Trends in Cognitive Sciences* 24 (2020), pp. 814–825.

[26] J. F. Storm et al. "An integrative, multiscale view on neural theories of consciousness". In: *Neuron* 112.10 (2024), pp. 1531–1552. DOI: 10.1016/j.neuron.2024.02.004.

[27] F. Benitez, C. Pennartz, and W. Senn. *The conductor model of consciousness, our neuromorphic twins, and the human-AI deal*. 2023. DOI: 10.31234/osf.io/gbzd6. URL: https://doi.org/10.31234/osf.io/gbzd6.

[28] N. Cortes et al. "The pulvinar as a hub of visual processing and cortical integration". In: *Trends in Neurosciences* (2023), S0166223623002709. DOI: 10.1016/j.tins.2023.11.008.

[29] U. Klatzmann et al. *A connectome-based model of conscious access in monkey cortex*. Preprint. 2022. doi: 10.1101/2022.02.20.481230.

[30] H. C. Lau. "A higher order Bayesian decision theory of consciousness". In: *Progress in Brain Research*. Vol. 168. Elsevier, 2007, pp. 35–48. DOI: 10.1016/S0079-6123(07)68004-2.

[31] V. A. F. Lamme. "Towards a true neural stance on consciousness". In: *Trends in Cognitive Sciences* 10.11 (2006), pp. 494–501. DOI: 10.1016/j.tics.2006.09.001.

[32] B. R. Munn et al. "A thalamocortical substrate for integrated information via critical synchronous bursting". In: *Proceedings of the National Academy of Sciences* 120.46 (2023), e2308670120. DOI: 10.1073/pnas.2308670120.

[33] K.D. Harris and G.M.G. Shepherd. "The neocortical circuit: themes and variations". In: *Nature Neuroscience* 18 (2015), pp. 170–181.

[34] M. Quiquempoix et al. "Layer 2/3 Pyramidal Neurons Control the Gain of Cortical Output". In: *Cell Reports* 24.11 (2018), 2799–2807.e4.

[35] A.B. Barron and C. Klein. "What Insects Can Tell Us about the Origins of Consciousness". In: *Proceedings of the National Academy of Sciences* 113 (2016), pp. 4900–4908.

[36] T. Bayne. "Unity of consciousness". In: *Scholarpedia Journal* 4 (2009), p. 7414.

[37] B. Merker. "Consciousness without a Cerebral Cortex: A Challenge for Neuroscience and Medicine". In: *Behavioral and Brain Sciences* 30 (2007), pp. 63–81.

[38] J.M. Shine. "Adaptively Navigating Affordance Landscapes: How Interactions between the Superior Colliculus and Thalamus Coordinate Complex, Adaptive Behaviour". In: *Neuroscience and Biobehavioral Reviews* 143 (2022), p. 104921.

[39] A. Canales-Johnson et al. "Dissociable Neural Information Dynamics of Perceptual Integration and Differentiation during Bistable Perception". In: *Cerebral Cortex* 30.8 (2020), pp. 4563–4580.

[40] M. Oizumi, L. Albantakis, and G. Tononi. "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0". In: *PLoS Computational Biology* 10 (2014), e1003588.
[41] G. Tononi and G.M. Edelman. "Consciousness and Complexity". In: *Science* 282 (1998), pp. 1846–1851.
[42] T. Bachmann, M. Suzuki, and J. Aru. "Dendritic integration theory: A thalamo-cortical theory of state and content of consciousness". In: *Philosophical Inquiries* 1 (2020), 10.33735/phimisci.2020.ii.52.
[43] W. James. *The Principles of Psychology (Vol. 1&2)*. Good Press, 2023.
[44] G. Tononi and C. Koch. "Consciousness: here, there and everywhere?" In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 370 (2015). DOI: 10.1098/rstb.2014.0167.
[45] T. Bayne. *The Unity of Consciousness*. Oxford University Press, 2012.
[46] D. Keller, C. Erö, and H. Markram. "Cell Densities in the Mouse Brain: A Systematic Review". In: *Frontiers in Neuroanatomy* 12 (2018), p. 83.
[47] C.S. von Bartheld, J. Bahney, and S. Herculano-Houzel. "The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting". In: *Journal of Comparative Neurology* 524 (2016), pp. 3865–3895.
[48] E. J. Müller, B. R. Munn, and J. M. Shine. "Diffuse neural coupling mediates complex network dynamics through the formation of quasi-critical brain states". In: *Nature Communications* 11.1 (2020), p. 6337.
[49] J.M. Shine et al. "The Impact of the Human Thalamus on Brain-Wide Information Processing". In: *Nature Reviews Neuroscience* 24 (2023), pp. 416–430.
[50] M.M. Halassa and Y.B. Saalmann. "The Mediodorsal Thalamus and Decision-Making". In: *The Cerebral Cortex and Thalamus*. Ed. by A.S.M.S.W. Martin Usrey. Oxford University Press, 2023.
[51] K. Hwang et al. "Thalamocortical Contributions to Cognitive Task Activity". In: *eLife* 11 (2022), e81282.
[52] J.M. Shine et al. "The Low-Dimensional Neural Architecture of Cognitive Complexity Is Related to Activity in Medial Thalamic Nuclei". In: *Neuron* 104 (2019), 849–855.e3.
[53] E. J. Müller et al. "Core and Matrix Thalamic Sub-Populations Relate to Spatio-Temporal Cortical Connectivity Gradients". In: *NeuroImage* 222 (2020), p. 117224.
[54] A.F. Sadikot and V.V. Rymar. "The primate centromedianparafascicular complex: Anatomical organization with a note on neuromodulation". In: *Brain Research Bulletin* 78 (2009), pp. 122–130.
[55] A.M. Ni, S. Ray, and J.H.R. Maunsell. "Tuned normalization explains the size of attention modulations". In: *Neuron* 73 (2012), pp. 803–813.
[56] D. Pinault. "The Thalamic Reticular Nucleus: Structure, Function and Concept". In: *Brain Research Reviews* 46 (2004), pp. 1–31.
[57] J.H. Reynolds and D.J. Heeger. "The Normalization Model of Attention". In: *Neuron* 61 (2009), pp. 168–185.
[58] G. Hesselmann. *Transitions Between Consciousness and Unconsciousness*. Routledge, 2019.
[59] M.A. Sommer and R.H. Wurtz. "Brain circuits for the internal monitoring of movements". In: *Annual Review of Neuroscience* 31 (2008), pp. 317–338.
[60] J.R. Duhamel, C.L. Colby, and M.E. Goldberg. "The updating of the representation of visual space in parietal cortex by intended eye movements". In: *Science* 255 (1992), pp. 90–92.

[61] M.A. Sommer and R.H. Wurtz. "Influence of the thalamus on spatial visual processing in frontal cortex". In: *Nature* 444 (2006), pp. 374–377.
[62] Y. Watanabe, K. Takeda, and S. Funahashi. "Population vector analysis of primate mediodorsal thalamic activity during oculomotor delayed-response performance". In: *Cerebral Cortex* 19 (2009), pp. 1313–1321.
[63] J. Cavanaugh, K. McAlonan, and R.H. Wurtz. "Organization of Corollary Discharge Neurons in Monkey Medial Dorsal Thalamus". In: *The Journal of Neuroscience* 40 (2020), pp. 6367–6378.
[64] M.A. Sommer and R.H. Wurtz. "What the brain stem tells the frontal cortex. I. Oculomotor signals sent from superior colliculus to frontal eye field via mediodorsal thalamus". In: *Journal of Neurophysiology* 91 (2004), pp. 1381–1402.
[65] J. Cavanaugh et al. "Saccadic Corollary Discharge Underlies Stable Visual Perception". In: *The Journal of Neuroscience* 36 (2016), pp. 31–42.
[66] A.S. Mitchell and S. Chakraborty. "What Does the Mediodorsal Thalamus Do?" In: *Frontiers in Systems Neuroscience* 7 (2013). DOI: 10.3389/fnsys.2013.00037.
[67] J. Peräkylä et al. "Causal Evidence from Humans for the Role of Mediodorsal Nucleus of the Thalamus in Working Memory". In: *Journal of Cognitive Neuroscience* 29 (2017), pp. 2090–2102.
[68] Y. Watanabe and S. Funahashi. "Neuronal Activity Throughout the Primate Mediodorsal Nucleus of the Thalamus During Oculomotor Delayed-Responses. I. Cue-, Delay-, and Response-Period Activity". In: *Journal of Neurophysiology* 92 (2004), pp. 1738–1755.
[69] M. Wolff and M.M. Halassa. "The mediodorsal thalamus in executive control". In: *Neuron* (2024). DOI: 10.1016/j.neuron.2024.01.002.
[70] D. Soto, T. Mäntylä, and J. Silvanto. "Working memory without consciousness". In: *Current Biology* 21 (2011), R912–R913.
[71] D. Soto and J. Silvanto. "Is conscious awareness needed for all working memory processes?" In: *Neuroscience of Consciousness* 2016 (2016), niw009.
[72] D. Trübutschek et al. "A theory of working memory without consciousness or sustained activity". In: *eLife* 6 (2017). DOI: 10.7554/eLife.23871.
[73] X.-J. Wang. "50 years of mnemonic persistent activity: quo vadis?" In: *Trends in Neurosciences* 44 (2021), pp. 888–901.
[74] S.S. Bolkan et al. "Thalamic projections sustain prefrontal activity during working memory maintenance". In: *Nature Neuroscience* 20 (2017), pp. 987–996.
[75] L.I. Schmitt et al. "Thalamic Amplification of Cortical Connectivity Sustains Attentional Control". In: *Nature* 545 (2017), pp. 219–223.
[76] J.M. Fuster and G.E. Alexander. "Neuron activity related to short-term memory". In: *Science* 173 (1971), pp. 652–654.
[77] J.M. Phillips et al. *Primate thalamic nuclei select abstract rules and shape prefrontal dynamics*. bioRxiv. 2024.
[78] D. Rigney. *How Advantage Begets Further Advantage*. Columbia University Press, 2010.
[79] C.J. Whyte and R. Smith. "The predictive global neuronal workspace: A formal active inference model of visual consciousness". In: *Progress in Neurobiology* 199 (2021), p. 101918.