Spring 5-15-2020

Applications of Nonlinear Dynamics, and Critical Phenomena to Measure Neural Populations using Inputs to Single Neurons

James Kenneth Johnson
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Physics

Dissertation Examination Committee:
Ralf Wessel, Chair
Anders Carlsson
Zohar Nussinov
Baranidharan Raman
Mikhail Tikhonov

Applications of Nonlinear Dynamics, and Critical Phenomena to Measure Neural Populations using Inputs to Single Neurons

by

James Kenneth Johnson

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2020
St. Louis, Missouri

Table of Contents

List of Figures v
List of Tables vii
Acknowledgments viii
Abstract x
Chapter 1: Introduction to Intracellular Investigations of Population Neural Code 1
1.1 Cellular Factors Influencing Neural Activity 2
1.2 Advantages of Whole Cell Recordings 7
1.3 Assessing the potential of whole cell recordings for facilitating neural code investigations 8
References 12
Chapter 2: Single-cell membrane potential fluctuations evince network scale-freeness and quasicriticality 15
2.1 Introduction 16
2.2 Methods 19
2.2.1 Surgery and Visual Cortex 19
2.2.2 Intracellular Recordings 21
2.2.3 Extracellular Recordings 22
2.2.4 Set-wise Comparisons 23
2.2.5 Random Surrogate Testing 24
2.2.6 Neuronal Avalanche Analysis 25
2.2.7 Model Simulations 34
2.2.8 Data and Software Accessibility 38
2.3 Results 38
2.3.1 Membrane Potential Fluctuations Reveal Signatures of Critical Point Dynamics 41
2.3.2 Membrane Potential Fluctuations are Consistent with Avalanches from Previously Obtained Microelectrode Array LFP Recordings 46
2.3.3 The Single-Neuron Estimate of Network Dynamics is Optimized at the Network Critical Point 51
2.3.4 The Predicted Scaling Relation Exponent is More Stable than Avalanche Size or Duration Exponents 63
2.3.5 Non-Linearity and Temporal Characteristics such as High-Order Correlation, Proper Combination of Synaptic Events, and Signal Time-Scale are Required to Reproduce Network Measures from Single-Electrode Recordings 66

2.3.6 Negative Fluctuations of LFP Disagree with V_m and MEA Results and are Inconsistent with Avalanches in Critical Systems 67
2.3.7 Stochastic Surrogates are Distinguishable from V_m or MEA Results, Reveal Importance of Non-Linear Filtering 73
2.3.8 Excitatory and Inhibitory Synaptic Activity are both Required for V_m Fluctuations to Match MEA Avalanches 76
2.4 Discussion 79
2.4.1 Neuronal Avalanche and Neural Input Fluctuation Similarity is Captured by a Critical Recurrent Coarse-Graining Network 81
2.4.2 Conclusion 84
References 86
Chapter 3: Stimulus tuned dynamical trajectories underlie synaptically driven transmembrane signals in visual cortex 99
3.1 Introduction: 100
3.2 Methods and Materials 103
3.2.1 Animals: Surgery and Electrophysiological Recording 104
3.2.2 Electrophysiology 105
3.2.3 Visual Stimulation 108
3.2.4 Tests and measures 108
3.2.5 Characterizing Recorded Responses with Deflection and Definition of Epoch 110
3.2.6 Single Neuron Modeling 113
3.2.7 Dimensionality Expansion 117
3.2.8 Maximum Likelihood Estimation 118
3.2.9 Reliability Tuning Curves 120
3.2.10 Genetic Modification of SINDy 122
3.2.11 Classifier Objective Function 126
3.2.12 Goodness of Fit Objective Function 127
3.2.13 Ensemble Classification and Out-of-Sample Generalization 128
3.2.14 Hyperparameter Optimization 129
3.2.15 Integration of Ξ Matrices 132
3.2.16 Linear Stability Analysis 133
3.2.17 Data and Software Availability 134
3.3 Results 135
3.3.1 Dimensionality Expansion Captures Dynamically Rich Neural Trajectories from Single Neurons 135

3.3.2 Dynamical Discrimination Reveals Smaller Changes in Stimulus Characteristics than Firing Rate Substitutes Reveal 138
3.3.3 Dynamical Discrimination is Linked to Distinct and Reliable Presynaptic Population Dynamics 142
3.3.4 Dynamical Discrimination is Corroborated by Biologically Plausible Modeling 146
3.4 Discussion 151
References 156
Chapter 4: Summary and Future Work 162
4.1 The Implications for Scale-Freeness and Criticality, Neural Code, and Data Analysis in Neuroscience 164
4.1.1 Aim One: Expand and Clarify the Study of Scale-Freeness and Criticality in Neural Systems 164
4.1.2 Aim Two: Articulate Intrinsic Representations of Neural Dynamics and Neural Code 167
4.1.3 Aim Three: Demonstrate the Appropriate Use of Machine-Learning to Test Scientific Hypotheses 172
4.2 Future directions 175
4.2.1 Expanding Neural Code and Criticality Comparisons 176
4.2.2 Improving and Extending Dynamical Discrimination Methods 178
4.2.3 Critical Averaging Robustness and Relationship to Homeostasis and General Computation 181
4.3 Conclusion 186
References 187
Appendix 192
A.1 Dynamical Discrimination as Timeseries Classifier: Performance and Additional Insights 192
A.2 Hyperparameter Optimization Reveals Epoch Dependence of Dynamical Discrimination 203
A.3 Analysis of Dynamical Stability Underscores an Attractor Dynamics Interpretation 211
A.4 Maximum Likelihood Estimation of Stimulus Reveals Dimensionality Expansion is Not Sufficient 218
References 221

List of Figures

Figure 2.1 | Will fluctuations in somatic membrane potential and comparable signals reflect the scale-free nature of neuronal avalanches from microelectrode array data? 19

Figure 2.2 | Membrane potential fluctuations reveal signatures of critical point dynamics. 40

Figure 2.3 | Membrane potential fluctuations are consistent with avalanches from previously obtained microelectrode array data. 48

Figure 2.4 | The single-neuron estimate of network dynamics is optimized at the network critical point. 50

Figure 2.5 | Inputs to a neuron stochastically estimate firing of its presynaptic pool in this critical branching model. 57

Figure 2.6 | Comparison to surrogate signals reveals the importance of non-linearity and temporal characteristics such as high-order correlation, proper combination of synaptic events, and signal time-scale. 68

Figure 2.7 | Plausible alternative signals fail to demonstrate consistency with criticality. 70

Figure 3.1 | The context of a dynamical trajectory is represented in governing equation parameters. 103

Figure 3.2 | Time-delay embedding of intracellular recordings reveals varied dynamical trajectories and equations are fitted to them. 137

Figure 3.3 | Discrimination performance of multiple methods across categories and compared with deflection for individual cells. 141

Figure 3.4 | Single trial discriminability depends on dynamical states associated with stimulus selectivity. 145

Figure 3.5 | Modeling tests confounding factors for dynamical discrimination and matches experimental results. 150

Figure A.1 | Deflection is useful as a binary discriminator in accordance with stimulus selectivity theory. 202

Figure A.2 | Comparisons for hyperparameter optimization part one: weak regularization, SVD compared with ICA along with epoch and dimension options. 207

Figure A.3 | Comparisons for hyperparameter optimization part two: varied regularization and dimensionality reduction along with epoch and dimension options. 208

Figure A.4 | Comparisons for hyperparameter optimization part 3: moderate regularization, four categories showing epoch and dimension dependence. 209

Figure A.5 | Comparisons for hyperparameter optimization part four: moderate regularization, four categories showing epoch and dimension dependence. 210

Figure A.6 | Comparisons for hyperparameter optimization part five, varied strong regularization along with epoch and dimension options. 211
Figure A.7 | Scatter plots summarizing differences between stimulus conditions according to various indicators of bifurcations. 212
Figure A.8 | Scatter plots showing how measures of noise and fluctuation size relate to deflection and reliability. 216
Figure A.9 | Trajectories vary by stimulus and occupy different but non-separable regions of state space. 220

List of Tables

Table A.1 | Summary of results for all algorithms and all categories. 198
Table A.2 | Binary discrimination data. 203

Acknowledgments

The work in this dissertation would not have been possible without the hard work of Ralf Wessel and 夏 (Xia) 霁 (Ji) of the Washington University in St Louis Department of Physics, Nathaniel C. Wright now at Georgia Institute of Technology, Dept. of Biomedical Engineering, 耿 (Geng) 松源 (Songyuan), Maximilian Wolfgang Hoffman of the Washington University School of Medicine, Hillel Adesnik of the University of California at Berkeley Department of Molecular and Cell Biology. These individuals collected the data, helped with data analysis and offered advisement. They are the co-authors of the publications resulting from this work.

This research was supported by a Whitehall Foundation grant (no. 20121221; R. Wessel) and a National Science Foundation Collaborative Research in Computational Neuroscience grant (no. 1308159; R. Wessel).

James Kenneth Johnson

Washington University in St. Louis

May 2020

Dedicated to my wife, Katharine, and to my darling daughter Annabelle.

ABSTRACT

Applications of Nonlinear Dynamics and Critical Phenomena to

Measure Neural Populations using Inputs to Single Neurons

by

James Kenneth Johnson

Doctor of Philosophy in Physics

Washington University in St. Louis, 2020

Professor Ralf Wessel, Chair

A compelling vision for the future of neuroscience is the ability to sense neural activity throughout the bulk of the brain with exquisite resolution. Popular visions usually include intricate electrode technology intruding into the neuropil, meandering along nerve tracts, and sensing the whole brain. These popular visions stem from the belief that we must always have an outsider's perspective of neural activity. According to this belief the closest thing neuroscientists can achieve to an insider's perspective is to shadow every neuron (or almost every neuron) with an electrical or optical recording device. Yet, the brain naturally has an expansive sensor network. The brain already aggregates and organizes neural activity according to computational function. The brain does this through the operation of single neurons, which have arrays of many dendrites to process inputs arriving from far and wide. These processed inputs are concentrated at the soma of the neuron where they drive rich dynamics, and where the neuron translates these inputs into outputs. One of the most venerable methods in neuroscience, the patch-clamp intracellular recording technique, can record these rich input driven dynamics. Neuroscience has

long held the goal of patching into the full network dynamics with patch-clamp, but it is difficult to reconstruct network dynamics information. Fortunately, the neural criticality hypothesis provides a justification for expecting to find network dynamics information, and the modern field of nonlinear dynamics provides tools for reconstructing full dynamics from scant information. The neural criticality hypothesis is the idea that the brain can exhibit phase transitions, but tunes itself to sit at a point (called a "critical point") between two phases where the most dynamical complexity arises. One of the key phenomena of critical systems is "scale-freeness" which is widely observed in the brain. One implication of Scale-freeness is that some statistics are always the same whether observed at very small scales or very large. For critical phenomena scale-freeness is both extensive and precise, if scale-freeness is limited in a system then it suggests that system is not a critical system. We adopt condensed matter physics' rigorous standards for experimentally identifying critical systems. We show that we can meet these standards with long intracellular recordings. We also show that our findings agree with large scale population recordings. After establishing this proof-of-concept, we then use new methods for modeling nonlinear dynamical systems to extract small details about visual stimulus from short intracellular recordings. These details were too small to be reliably detected in the output of neurons. We use models of nonlinear dynamics because of their relationship to a neural coding paradigm: Attractor network theory. Thus, we also have novel evidence of dynamical attractor based neural code in primary visual cortex. Therefore, we have advanced both the neural criticality hypothesis and the attractor network theory of neural coding while demonstrating that we can patch into in-situ neural communication networks and get information that previously required electrode arrays or other population recording methods.

Chapter 1:
Introduction to Intracellular Investigations of Population Neural Code

Throughout the last thirty years of innovation, reported breakthroughs, and real-world applications the study of neural code has continued to face an epistemological challenge familiar to science (Stanley, 2013). The term neural code is very general and refers to a hitherto undiscovered set of rules relating arbitrary behaviors and stimuli to a well-defined (though likely infinite) set of spatiotemporal patterns of brain activity for any organism. Given the profound complexity and fragility of vertebrate nervous systems, observing, characterizing, and validating any putative neural code is one of the most difficult problems in science. The core dilemma of this, and many research programs is: when an experimenter finds patterns of neural activity that allow them to predict external observables (e.g. stimulus, behavior), how can they know whether the patterns are causal or coincidental? To make things simpler researchers focus on limited versions of neural code. Plausible limited neural code concepts such as rate-coding have been revised and limited as new technologies and experiments provide counterexamples or alternatives. For the rate coding example: each neuron exhibits selectivity (i.e. they "stand for something") and thus fire more action-potentials when they become relevant to an organism objective (Georgopoulos et al., 1986; Bialek et al., 1991; Butts and Goldman, 2006). There are newer discoveries about the sensitivity to the precise timing of action potentials (Desbordes et al., 2008) and the finding that selectivity is context dependent, a neuron responds to visual

stimuli differently when an animal is locomoting than when still (Vinck et al., 2015). These new findings proved that although rate coding is very successful as a paradigm for predicting external observables from neural activity, firing rate may not be a fundamental quantity, but rather a side-effect of or a single element of a more fundamental computational paradigm. A core reason for this epistemological dilemma is that experimenters are forced to observe neural activity with an outsider's perspective that misses structural and contextual information. The lack of ability to gain an insider's perspective on brain operations and the combinatorial explosion of possible patterns that can be observed in neural activity is one of the principal reasons cracking the neural code has been nominated as the hardest problem in science (Horgan, n.d.).

The simultaneous activity of many neurons is primarily recorded using two methods, electrode arrays which sense electrotonic fluctuations (Spira and Hai, 2013), or imaging methods that observe changes in the fluorescence of molecules that label anatomical elements and are modulated by electrotonic variables (Yang and Yuste, 2017). Neither method is privy to the connections between neurons, they primarily sense activity at the point of action potential generation or are not localizable at the resolution needed to control for the myriad of factors that actively participate in a healthy brain. These factors can influence the timing of action potentials, communicate context, or alter the way that neurons pass information among each other.

1.1 Cellular Factors Influencing Neural Activity

We will quickly recount a few functional and anatomical details (Squire et al., 2012) before articulating how they thwart efforts to define neural code. The physical shape of a single neuron is typically modeled as something like a tree, with an exquisite dendritic arborization collecting inputs and a meandering tap-root-like axonal process sending information far and wide. The fastest and most easily observed carrier of information is the action-potential which is a self-propagating electrotonic impulse. These can be generated at a few key sites in the dendritic arborization of some neuron types, but historically the most important site for action potential generation has been near the cell-nucleus (i.e. the soma), at the "axon-hillock" which would be analogous to the root-bulge at the base of a tree. Action potentials must travel along narrow fibers and they rely on ion-channels embedded in those fibers to create or eliminate a separation of charges from inside the fiber to outside. Because this separation of charges sets up a fast-changing electric field, electrodynamic interactions play a large role in governing the propagation of action-potentials. However, because propagation also relies on ion channels, metabolic and biomolecule signaling factors play a role. Perhaps the most crucial part of action-potential propagation is its encounter with synapses. As an action potential propagates through the axonal processes the wave will pass across sites where the end-output organs of neurons are located. At a synapse the membrane of an axonal fiber comes very close to the membrane of a dendritic fiber. Sometimes synapses form between two dendritic fibers, or two axonal fibers but axon to dendrite synapses are the canonical model. A synapse also has a distinct shape, like a cup and saucer where a protrusion of the axonal membrane, called the presynaptic terminal, is convex (the cup) and a protrusion of the dendritic membrane, called the postsynaptic terminal is concave (the saucer). Membranes do not

touch but are separated by a narrow “synaptic cleft”. There are other, highly variable, details about the shapes of synaptic structures which are less relevant to our recitation of factors complicating experimental observations of neural code. More relevant are electronic and biomolecule signaling features. Ion channels are especially concentrated at synapses. For synapses called “gap-junctions”, the passing wave of electric-potential is felt on both sides and is enough to trigger the ion-channels on the postsynaptic side to induce a new wave in the dendritic fibers. In “chemical synapses” the passing wave induces the release of signaling molecules to trigger a similar wave. In both cases, this post-synaptic wave is not self-propagating and will decay to negligibility unless it constructively interferes with other post-synaptic waves or contributes to the generation of a dendritic action potential. Some kinds of synapses are inhibitory, others are excitatory. The waves induced by excitatory synapses increase the likelihood of somatic action potential generation and are usually associated with presynaptic neurons whose axonal processes feature almost exclusively excitatory synapses. An inhibitory synapse produces waves which decrease the likelihood of action potential generation and are usually associated with presynaptic neurons whose axonal processes feature almost exclusively inhibitory synapses. This brief anatomical description highlights the opportunities for mechanisms that electrode arrays and fluorescence imaging methods may miss.

There are three mechanisms relevant to the above details that are missed by most population recording methods, and which we aim to include in neural code investigations by using whole-cell recordings. The first is recurrent dynamics within the neural network being partially observed. Because axons can project far and wide a large amount of any

brain system occurs outside the imaging field or bounds of an electrode array. Axons can project from one brain hemisphere to the other, to neighboring neurons, or just across layers. Therefore, a single functional subgraph (a collection of interconnected neurons cooperatively performing an operation) can extend beyond any reach of current cellular resolution imaging and array recording methods (Levina and Priesemann, 2017). Even for the neurons which are in the imaging field, the synapses and fibers are almost always resolved too poorly to create a map of interconnections using anatomical evidence. Though this is changing (Viventi et al., 2011; Steinmetz et al., 2018). Thus, limited coverage and poor resolution together mean that electrode array recordings and fluorescence imagine methods do not give enough information for an experimenter to efficiently discover natural groupings of neurons and look for neural code within only those groupings and the complete grouping.

The second mechanism which is important but not well resolved in electrode arrays or imaging fields are biomolecule signaling effects. The structure of synapses provides imperfect isolation, consequently enabling the important process of “neuromodulation” (Marder, 2012). Regions of neural tissue can show altered function by any process altering the concentrations of neurotransmitters available at the synaptic cleft and other molecules facilitating signaling and/or cell metabolism. These effects can be selective for certain synapse types or localized to certain regions. Thus, for the largest imaging fields and arrays, these effects can be heterogeneously distributed, affecting some neurons under observation more than others. Since effects on action potentials are observed, but not the biomolecule signaling itself, neuromodulation and related phenomena are difficult to

properly account for. Neuromodulation and other biomolecule signaling events are much slower than action potential generation and propagation but may still play a role in computation, especially with regards to altering apparent connectivity (Elsegai et al., 2015). Furthermore, neurons have complex internal signaling networks which can alter the behavior of their membranes. These dynamics are unobserved and unique for each cell.

The third mechanism is dynamical complexity. The proximate cause for action potential generation is when a neuron depolarizes (the separation of charges across membrane decreases) to the point that it triggers a reaction in certain ion channels. Put more simply, the proximate cause of action potential generation is when transmembrane potential crosses a threshold. However, a feature of the Hodgkin-Huxley model of action potential generation is that the mechanisms governing separation of charges has momentum. If a neuron is held in a hyperpolarized state and is rapidly released the restoring mechanism can overshoot the point of equilibrium and reach the point of action potential generation. In this situation, called "rebound spiking", the (Izhikevich, 2007) ultimate cause can be something typically associated with inhibition and occurs well before it crosses the action potential generation threshold. The Hodgkin-Huxley model is quite simple and does not include neuromodulatory or intracellular signaling effects. Therefore, other ultimate causes for action potential generation may occur in the tissue under observation, but not be observable using current methods. This is because electrode arrays and fluorescence imaging methods are primarily designed to record action potentials and aren't intended to extract the upstream population spiking or biomolecular signaling events which may ultimately cause action potentials and influence neural code. These three factors,

incomplete coverage of neural groups and unknown connectivity, unobserved modulatory mechanisms altering activity and connectivity, and dynamical complexity all have something in common. Their effects all ultimately govern the electronic fluctuations that wash over each neuron's cell body to reach the axon hillock.

1.2 Advantages of Whole Cell Recordings

A single whole-cell recording of membrane potential yields a summative record of these factors, while recordings of transmembrane current can separate them to some extent. The signal recorded from a patch electrode located at a neuron's cell body is not unlike recording the sum of neurons under observation with an electrode array or calcium imaging field. However, it is also congruent with the intrinsic neural network in four important ways. First, it is no longer a presumption to say that a neuron has a well-defined effect on brain operations. Since an array or imaging field can happen to sit at the boundary of two functional subgraphs, this a presumption with those methods. A single neuron is a single computational element, the experimenter may not know its well-defined function or how reliably the neuron performs it, but they may trust that it has one.

Second, the superposition of excitatory and inhibitory postsynaptic potentials (postsynaptic waves) that an experimenter observes as fluctuations in whole-cell membrane potential recordings amounts to a way of subsampling neurons. Imaging fields and electrode arrays capture the individual activity of many neurons, but analysis frequently involves combining them in some manner. With whole-cell membrane potential

recordings they come pre-combined but can be separated with voltage-clamp recording modalities or with advanced computational data analysis. Many of the same conclusions about populations can be reached. Even better, we can count on the fact that the subsample of neurons summarized with a whole-cell recording is also a single intrinsically defined anatomical group that has a well-defined function.

Third, because we have only one cell under observation then we have only one set of neuromodulatory and intracellular signaling effects. While still not directly observed we no longer have the complication that comes from having different dimensions subjected to different factors.

Fourth, any events that ultimately cause action potentials are reflected in, though not easily recovered from, the fluctuations of transmembrane current or membrane potential. Action potential triggering events may occur among the inputs or within the components of the dendritic arborization.

These four elements of whole-cell recordings motivates careful comparison of analysis on whole-cell recordings to similar analyses performed on recordings of populations of neurons. By making this comparison and carefully choosing what type of neuron to record from we can gauge whether neural code findings are robust to more naturalistic observation methods.

1.3 Assessing the potential of whole cell recordings for facilitating neural code investigations

We assessed the potential of using whole-cell recordings to co-opt single neurons as useful naturally in-situ observers of brain operations by making two intensive comparisons to prior knowledge about brain operations obtained from recordings of neural populations. The first, covered in chapter two is a powerful test of the precision with which we can measure subtle properties of neural populations. What's more, the investigation of chapter two is rooted in a theoretical principle which also lends plausibility to measuring neural populations through single intracellular recordings. This theory, neural criticality (Plenz and Niebur, 2014), is the proposal that the brain operates at a fine point of balance, known as the critical point of a phase transition. This point optimizes many important properties relevant to information processing but is also a possible explanation for the widespread observation of scale freeness in brain signals. Scale-freeness is the finding that some statistics of brain operation are the same no matter what scale they are recorded from (He, 2014). Criticality provides a set of statistics predicted to be scale free but also requires that the match across scales be very precise; deviation is inconsistent with criticality (Taylor et al., 2013; Hartley et al., 2014). This is elaborated on more completely in chapter two, but it is important to understand that the concept of scale-freeness provides a broad justification for expecting to find evidence of a population neural code in intracellular recordings. What's more, demonstrating consistency with criticality is a difficult challenge and thus a useful proof of concept. We found that the geometry of membrane potential fluctuations provided close agreement to the geometry of spurts of population activity called neuronal avalanches. This agreement provided evidence for criticality at the smallest possible scale of neural activity observation and justified further explorations of the ability to measure neural populations with single whole-cell recordings.

The second intensive comparison to prior knowledge about brain operations obtained from recordings of neural populations aims to find a version of neural code directly in whole-cell recordings and is covered in chapter three. The basis of this code is provided by the attractor network paradigm for neural computation (Amit and Amit, 1992; Wu and Amari, 2005; Wu et al., 2008). Identification of this kind of neural code is enabled by dimensionality expansion and the automated discovery of differential equations (Brunton et al., 2016); both of which are novel methods to neuroscience. The attractor network paradigm takes as its premise that brain operation is stereotyped to a certain extent. If a memory is recalled twice, or a motion made twice, or image seen twice, then a pattern of neural activity is approximately repeated twice in each instance. However, it allows for a great deal of abstraction. With the attractor network paradigm, it is possible that none of the same neurons are active in each instance. It models high-dimensional population activity as an embedding of a moderate dimension dynamical system. It then associates regions or trajectories within this moderate dimension space with certain brain operations. Just as high dimensional neural data can be subjected to dimensionality reduction to find moderate dimension trajectories (Cunningham and Yu, 2014), one-dimensional whole cell recordings can be subjected to dimensionality expansion to obtain moderate dimension trajectories (Sauer et al., 1991). We characterized these moderate dimension trajectories with systems of ordinary differential equations (ODEs) to look for equation parameters that are different under different stimulus conditions and tried to predict what stimulus was presented concurrent with each recording. The neurons were in mouse primary visual cortex and are already stimulus tuned, the most and least preferred stimulus (which differ greatly) can be easily discriminated by looking at deflection from baseline. Thus, in order to

evince the discovery of neural population code from single neurons we must demonstrate the ability to discriminate finer changes in stimulus than can be recovered from deflection from baseline. It would also advance the case if the results recapitulated other findings about stimulus representation in neural data. We found that we could discriminate fine distinctions in drifting gratings by summarizing whole-cell recordings with fitted ODEs rather than by deflection from baseline. Furthermore, we found that we could discriminate orientation when fitting ODEs to excitation but not to inhibition. This recapitulates the finding that orientation information arrives in excitatory thalamocortical projections to mouse primary visual cortex (Sun et al., 2016). With these two findings we can establish that whole-cell recordings are amenable to analyses aiming to recover details about neural populations. Furthermore, by looking for findings conserved between population recordings and whole-cell recordings researchers can better constrain and corroborate theories about the neural code.

References

Amit DJ, Amit DJ (1992) Modeling Brain Function: The World of Attractor Neural Networks. Cambridge University Press. Available at: https://play.google.com/store/books/details?id=fvLYch1yQncC.

Bialek W, Rieke F, de Ruyter van Steveninck RR, Warland D (1991) Reading a neural code. Science 252:1854–1857 Available at: http://dx.doi.org/10.1126/science.2063199.

Brunton SL, Proctor JL, Kutz JN (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc Natl Acad Sci U S A 113:3932–3937 Available at: http://dx.doi.org/10.1073/pnas.1517384113.

Butts DA, Goldman MS (2006) Tuning curves, neuronal variability, and sensory coding. PLoS Biol 4:e92 Available at: http://dx.doi.org/10.1371/journal.pbio.0040092.

Cunningham JP, Yu BM (2014) Dimensionality reduction for large-scale neural recordings. Nat Neurosci 17:1500–1509 Available at: http://dx.doi.org/10.1038/nn.3776.

Desbordes G, Jin J, Weng C, Lesica NA, Stanley GB, Alonso J-M (2008) Timing precision in population coding of natural scenes in the early visual system. PLoS Biol 6:e324 Available at: http://dx.doi.org/10.1371/journal.pbio.0060324.

Elsegai H, Shiells H, Thiel M, Schelter B (2015) Network inference in the presence of latent confounders: the role of instantaneous causalities. J Neurosci Methods 245:91–106 Available at: http://dx.doi.org/10.1016/j.jneumeth.2015.02.015.

Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. Science 233:1416–1419 Available at: http://dx.doi.org/10.1126/science.3749885.

Hartley C, Taylor TJ, Kiss IZ, Farmer SF, Berthouze L (2014) Identification of Criticality in Neuronal Avalanches: II. A Theoretical and Empirical Investigation of the Driven Case. J Math Neurosci 4:9 Available at: http://dx.doi.org/10.1186/2190-8567-4-9.

He BJ (2014) Scale-free brain activity: past, present, and future. Trends Cogn Sci 18:480–487 Available at: http://dx.doi.org/10.1016/j.tics.2014.04.003.

Horgan J (n.d.) The Singularity and the Neural Code. Scientific American Blog Network Available at: https://blogs.scientificamerican.com/cross-check/the-singularity-and-the-neural-code/ [Accessed March 4, 2020].

Izhikevich EM (2007) Dynamical Systems in Neuroscience. MIT Press. Available at: https://play.google.com/store/books/details?id=kVjM6DFk-twC.

Levina A, Priesemann V (2017) Subsampling scaling. Nat Commun 8:15140 Available at: http://dx.doi.org/10.1038/ncomms15140.

Marder E (2012) Neuromodulation of neuronal circuits: back to the future. Neuron 76:1–11 Available at: http://dx.doi.org/10.1016/j.neuron.2012.09.010.

Plenz D, Niebur E (2014) Criticality in Neural Systems. John Wiley & Sons. Available at: https://play.google.com/store/books/details?id=r2diAwAAQBAJ.

Sauer T, Yorke JA, Casdagli M (1991) Embedology. J Stat Phys 65:579–616 Available at: https://doi.org/10.1007/BF01053745.

Spira ME, Hai A (2013) Multi-electrode array technologies for neuroscience and cardiology. Nat Nanotechnol 8:83–94 Available at: http://dx.doi.org/10.1038/nnano.2012.265.

Squire L, Berg D, Bloom FE, du Lac S, Ghosh A, Spitzer NC (2012) Fundamental Neuroscience. Academic Press. Available at: https://play.google.com/store/books/details?id=QGzJFu_NyzcC.

Stanley GB (2013) Reading and writing the neural code. Nat Neurosci 16:259–263 Available at: http://dx.doi.org/10.1038/nn.3330.

Steinmetz NA, Koch C, Harris KD, Carandini M (2018) Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. Curr Opin Neurobiol 50:92–100 Available at: http://dx.doi.org/10.1016/j.conb.2018.01.009.

Sun W, Tan Z, Mensh BD, Ji N (2016) Thalamus provides layer 4 of primary visual cortex with orientation- and direction-tuned inputs. Nat Neurosci 19:308–315 Available at: http://dx.doi.org/10.1038/nn.4196.

Taylor TJ, Hartley C, Simon PL, Kiss IZ, Berthouze L (2013) Identification of Criticality in Neuronal Avalanches: I. A Theoretical Investigation of the Non-driven Case. J Math Neurosci 3:5 Available at: http://dx.doi.org/10.1186/2190-8567-3-5.

Vinck M, Batista-Brito R, Knoblich U, Cardin JA (2015) Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. Neuron 86:740–754 Available at: http://dx.doi.org/10.1016/j.neuron.2015.03.028.

Viventi J et al. (2011) Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity in vivo. Nat Neurosci 14:1599–1605 Available at: http://dx.doi.org/10.1038/nn.2973.

Wu S, Amari S-I (2005) Computing with continuous attractors: stability and online aspects. Neural Comput 17:2215–2239 Available at: http://dx.doi.org/10.1162/0899766054615626.

Wu S, Hamaguchi K, Amari S-I (2008) Dynamics and computation of continuous attractors. Neural Comput 20:994–1025 Available at: http://dx.doi.org/10.1162/neco.2008.10-06-378.

Yang W, Yuste R (2017) In vivo imaging of neural activity. Nat Methods 14:349–359 Available at: http://dx.doi.org/10.1038/nmeth.4230.

Chapter 2:
Single-cell membrane potential fluctuations evince network scale-freeness and quasicriticality

What information single neurons receive about general neural circuit activity is a fundamental question for neuroscience. Somatic membrane potential fluctuations are driven by the convergence of synaptic inputs from a diverse cross-section of upstream neurons. Furthermore, neural activity is often scale-free implying that some measurements should be the same, whether taken at large or small scales. Together, convergence and scale-freeness support the hypothesis that single membrane potential recordings carry useful information about high-dimensional cortical activity. Conveniently, the theory of "critical branching networks" (one purported explanation for scale-freeness) provides testable predictions about scale-free measurements which are readily applied to membrane potential fluctuations. To investigate, we obtained whole-cell current clamp recordings of pyramidal neurons in visual cortex of turtles with unknown genders. We isolated fluctuations in membrane potential below the firing threshold and analyzed them by adapting the definition of "neuronal avalanches" (spurts of population spiking). The membrane potential fluctuations we analyzed were scale-free and consistent with critical branching. These findings recapitulated results from large-scale cortical population data obtained separately in complementary experiments using microelectrode arrays (previously published (Shew et al., 2015)). Simultaneously recorded single-unit local field

potential did not provide a good match, demonstrating the specific utility of membrane potential. Modeling shows that estimation of dynamical network properties from neuronal inputs is most accurate when networks are structured as critical branching networks. In conclusion, these findings extend evidence of critical phenomena while also establishing subthreshold pyramidal neuron membrane potential fluctuations as an informative gauge of high-dimensional cortical population activity.

2.1 Introduction

How do cortical population dynamics impact single neurons? What can we learn about cortical population dynamics from single neurons? These questions are central to neuroscience. Uncovering the functional significance of multiscale organization within cerebral cortex requires knowing the relationship between the dynamics of networks and individual neurons within them (Nunez et al., 2013).

For pyramidal neurons in the visual cortex, somatic spike generation is ambiguously related to presynaptic firing (Tsodyks and Markram, 1997; Brunel et al., 2014; Gatys et al., 2015; Stuart and Spruston, 2015; Moore et al., 2017). Such neurons pass spiking information to many postsynaptic neurons (Lee et al., 2016). However, a presynaptic pool with multifarious neighboring and distant neurons (Hellwig, 2000; Wertz et al., 2015) provides excitatory and inhibitory synaptic inputs throughout the soma and complex dendritic architecture (Magee, 2000; Larkum et al., 2008; Moore et al., 2017). Input propagation to the axon hillock has both active and passive features (London and Häusser,

2005), and the membrane potential (V_m) response is increasingly non-linear near the action potential threshold. Thus, such details of network propagation give membrane potential more utility than focusing solely on spiking.(Johnson et al., 2019)

Most computational neuroscientists use spiking data because spikes are "the currency of the brain" (Wolfe et al., 2010), and extracellular recording is straightforward compared to whole-cell recording. Yet, the paucity of single-neuron spiking (Shoham et al., 2006), and limited foreknowledge about connections (Helmstaedter, 2013) makes extracellular single-unit observation an impoverished means of studying neuronal circuits. In contrast, subthreshold V_m fluctuations contain rich information about the circuits containing each neuron (Sachidhanandam et al., 2013; Petersen, 2017). Integral to gaining a neuron's view of the brain is uncovering relationships between the statistics of V_m fluctuations and fluctuations of local spiking; then contrasting against other plausible one-dimensional signals.

We look for such relationships in the strict predictions and rigorous measurements of scale-freeness used to identify a fragile network connectivity pattern known as "critical branching". This pattern exhibits emergent properties valuable for information processing, such as higher susceptibility and dynamic range (Haldeman and Beggs, 2005; Beggs, 2008; Shew and Plenz, 2012; Shriki and Yellin, 2016; Timme et al., 2016), but omits some neuronal dynamics (Poil et al., 2008, 2012) without extension (Porta and Copelli, 2018). The pattern is as follows: on average over all neuronal avalanches (spiking above baseline (Friedman et al., 2012)), one spike leads to *exactly* one other spike. In most arbitrary

networks there is less or more than one; these are “subcritical” and “supercritical” respectively. Among the dazzling emergent properties of “criticality” are universality, self-similarity, and scale-free correlations (Stanley, 1999).

These are as follows: A “universality class” is a set of incongruous systems exhibiting identical statistics only at their “critical points”. “Self-similarity” includes fractal patterns and power-laws in geometrical analysis of avalanches (power-laws are “scale-invariant”, popularly called “scale-free”). Avalanches of any duration have identical average shapes after normalization (Shaukat and Thivierge, 2016). Avalanche areas grow with duration as another power-law (Sethna et al., 2001). However, observation methods must be consistent with event propagation (Priesemann et al., 2009; Yu et al., 2014; Levina and Priesemann, 2017). Additionally, pairwise correlation vs length or time are also power-laws (Chialvo, 2010) meaning any input has a nonzero chance of propagating forever or anywhere.

In summary, the theory of critical branching networks offers superb standards of comparison for three reasons: neuronal avalanche analysis applies to membrane potentials, offers promising insights, and makes precise predictions about fluctuation geometry. We study both V_m fluctuations and criticality with one simple question: Do V_m fluctuations match the scale-free statistics of cortical populations (Figure 2.1)?

To address this question, we simultaneously recorded somatic V_m from pyramidal neurons and local field potential (LFP) in visual cortex and performed avalanche analysis on

fluctuations. We found that subthreshold V_m fluctuation statistics match published microelectrode array (MEA) data. We used surrogate testing to show why negative LFP fluctuations don't match and modeling to demonstrate dependence on critical branching.

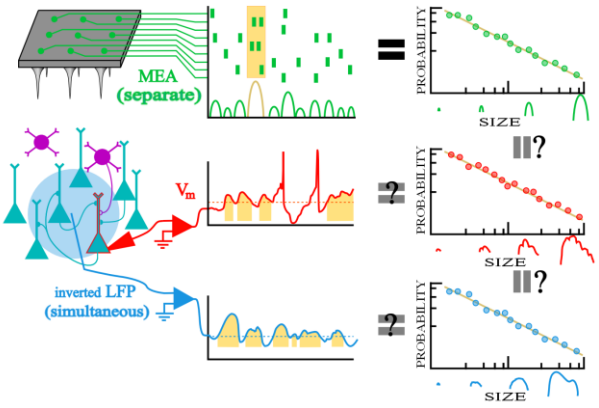

Figure 2.1 | Will fluctuations in somatic membrane potential and comparable signals reflect the scale-free nature of neuronal avalanches from microelectrode array data? A recurrent network with excitatory (teal) and inhibitory (purple) neurons is measured in three ways: microelectrode array (MEA)(green/upper), whole-cell recording (red/middle), LFP (blue/bottom). Neuronal avalanches (highlighted in gold) are inferred from the population raster and fluctuations are analyzed like avalanches for the V_m and inverted LFP signals. Neuronal avalanches are defined as spurts of activity with quiet periods between them for MEA or excursions above the 25th percentile for continuous non-zero data. The ultimate question is whether membrane potential fluctuations will recapitulate the entire neuronal avalanche analysis previously conducted on MEA data, including power-laws in size and duration as well as a universal avalanche shape. This is abridged in the right most column which illustrates power-law distributions.

2.2 Methods

2.2.1 Surgery and Visual Cortex

All procedures were approved by Washington University's Institutional Animal Care and Use Committees and conform to the guidelines of the National Institutes of Health on the Care and Use of Laboratory Animals. Fourteen adult red-eared sliders (*Trachemys scripta elegans*, 150-1000 g) were used for this study, their genders were not recorded. Turtles

were anesthetized with Propofol (2 mg Propofol/kg), then decapitated. Dissection proceeded as described previously (Saha et al., 2011; Crockett et al., 2015; Wright et al., 2017a).

To summarize, immediately after decapitation, the brain was excised from the skull, with right eye intact, and bathed in cold extracellular saline (in mM, 85 NaCl, 2 KCl, 2 $MgCl_2$*$6H_2O$, 20 Dextrose, 3 $CaCl_2$-$2H_2O$, 45 $NaHCO_3$). The dura was removed from the left cortex and right optic nerve, and the right eye hemisected to expose the retina. The rostral tip of the olfactory bulb was removed, exposing the ventricle that spans the olfactory bulb and cortex. A cut was made along the midline from the rostral end of the remaining olfactory bulb to the caudal end of the cortex. The preparation was then transferred to a perfusion chamber (Warner RC-27LD recording chamber mounted to PM-7D platform) and placed directly on a glass coverslip surrounded by Sylgard. A final cut was made to the cortex (orthogonal to the previous and stopping short of the border between medial and lateral cortex) allowing the cortex to be pinned flat, with ventricular surface exposed. Multiple perfusion lines delivered extracellular saline to the brain and retina in the recording chamber (adjusted to pH 7.4 at room temperature).

We used a phenomenological approach to identify the visual cortex, described previously (Shew et al., 2015). In brief, this region was centered on the anterior lateral cortex, in agreement with voltage-sensitive dye studies (Senseman and Robbins, 1999, 2002). Anatomical studies identify this as a region of cortex receiving projections from lateral geniculate nucleus (Mulligan and Ulinski, 1990). We further identified a region of neurons

as belonging to the visual cortex when the average LFP response to visual stimulation crossed a given threshold and patched within that neighborhood (radius of about 300 μm).

2.2.2 Intracellular Recordings

For whole-cell current clamp recordings, patch pipettes (4-8 MΩ) were pulled from borosilicate glass and filled with a standard electrode solution (in mM; 124 $KMeSO_4$, 2.3 $CaCl_2$-$2H_2O$, 1.2 $MgCl_2$, 10 HEPES, 5 EGTA) adjusted to pH 7.4 at room temperature. Cells were targeted for patching using a differential interference contrast microscope (Olympus). Membrane potential recordings were collected using an Axoclamp 900A amplifier, digitized by a data acquisition panel (National Instruments PCIe-6321), and recorded using a custom LabVIEW program (National Instruments), sampling at 10 kHz. As described in (Crockett et al., 2015; Wright and Wessel, 2017; Wright et al., 2017b, 2017a), before recording from a cell after initial patching current was injected to elicit spiking. This was also repeated intermittently between recording trials. Recording did not proceed if a cell spiked inconsistently (failure to spike, insufficient spike amplitude) in response to injected current, or exhibited extreme depolarization in response to small current injection amplitudes. If a clog or loss of seal was suggested by unusually erratic membrane potential short timescales current the current injection test was performed and upon failure, the affected recording was marked for exclusion from analysis. We excluded cells that did not display stable resting membrane potentials for long enough to gather enough avalanches. Up to 3 whole-cell recordings were made simultaneously. In total, we obtained recordings from 51 neurons from 14 turtles.

Recorded V_m fluctuations taken in the dark (no visual stimulation) were interpreted as ongoing activity. Such ongoing cortical activity was interrupted by visual stimulation of the retina with whole-field flashes and naturalistic movies as described previously (Wright and Wessel, 2017; Wright et al., 2017a, 2017b). An uninterrupted recording of ongoing activity lasted for 2 to 5 minutes. Periods of visual stimulation were too short and were too frequently interrupted by action potentials to yield the great number of avalanches which are required for rigorous power-law fitting.

A sine-wave removal algorithm was used to remove 60 Hz line noise. Action potentials in turtle cortical pyramidal neurons are relatively rare. An algorithm was used to detect spikes, the V_m recordings between spikes were extracted and filtered from 0 to 100 Hz. Membrane potential recordings were detrended by subtracting the 5^{th} percentile in a sliding 2 s window. The resulting signal was then shifted to have the same mean value as before subtraction. De-trending did not affect the size of membrane potential fluctuations (data not shown).

2.2.3 Extracellular Recordings

Extracellular recordings were achieved with tungsten microelectrodes (microprobes heat-treated tapered tip), with approximately 0.5 MΩ impedance. Electrodes were slowly advanced through tissue under visual guidance using a manipulator (Narishige), while monitoring for activity using custom acquisition software (National Instruments). The extracellular recording electrode was located within approximately 300 μm of patched neurons. Extracellular activity was collected using an A-M Systems Model 1800 amplifier,

band-pass filtered between 1 Hz and 20,000 Hz, digitized (NI PCIe-6231), and processed using custom software (National Instruments). Extracellular recordings were down-sampled to 10,000 Hz and then filtered (100 Hz low-pass), yielding the local field potential (LFP). The LFP was filtered and detrended as described above (see Intracellular Recordings), except that the mean of the entire signal was subtracted, and the signal was multiplied by -1 before it was detrended. This final inverted signal is commonly featured in literature as negative LFP or nLFP (Kelly et al., 2010; Kajikawa and Schroeder, 2011; Okun et al., 2015; Ness et al., 2016).

2.2.4 Set-wise Comparisons

In order to measure differences between sets of statistics we rely on three non-parametric measures. We use the MATLAB Statistics and Machine Learning Toolbox implementation of Fisher's exact test (Hammond et al., 2015). This lets us measure the effect size (Odds Ratio r_{OR}) and statistical significance (p value) of finding that consistency with criticality is more frequent or less frequent in an experimental group than a control group.

To quantify the similarity between the exponents measured in different sets of data we use the MATLAB Statistics and Machine Learning Toolbox implementations of the *exact* Wilcoxon rank sum test (Hammond et al., 2015) and the *exact* Wilcoxon signed rank test . In both cases effect size, r_{SDF} is measured by the simple difference formula (Kerby, 2014). The rank sum test is used when comparing non-simultaneous recordings, such as comparing MEA data with V_m data. The signed rank test is used when comparing data that

can be paired, such as V_m data to concurrent LFP. When comparing whether a dataset differs from a specific value, we can use the sign test.

The significance level is set at $p=0.05$ for all tests. Each set-wise comparison test stands alone as its own conclusion. None are combined to assess the significance of any effect across sets-of-sets. Thus, we are not making multiple comparisons and no corrections are warranted (Bender and Lange, 2001).

2.2.5 Random Surrogate Testing

It is possible that scale-free observations have an origin in independent random processes of a kind previously demonstrated (Touboul and Destexhe, 2017). To control for this, we phase-shuffled the V_m fluctuations using the amplitude adjusted Fourier transform (AAFT) algorithm (Theiler et al., 1992). This tests against the null hypothesis that a measure on a time series can be reproduced by performing a non-linear rescaling of a linear Gaussian process with the same autocorrelation (same Fourier amplitudes) as the original process. Phase information is randomized, which removes higher-order correlations but preserves the scale-free power-spectrum.

The AAFT tests only higher-order correlations, but a simpler algorithm tests against the null hypothesis that an *un-rescaled* linear Gaussian process with the same autocorrelation as the original process can produce the same results (Theiler et al., 1992). This is known as the Unwindowed Fourier Transform (UFT). Once we see what measures depend on the higher-order correlations with the AAFT we can use the UFT to see how measures depend

on the non-Gaussianity (non-linear rescaling) which is inherent to excitable membranes. Using the UFT alone would make it difficult to attribute whether statistically significant differences are due to the rescaling or to the higher-order correlations (Rapp et al., 1994).

We performed AAFT and UFT on each V_m time series once, and then compared how the two datasets performed on every metric used in this study. The datasets were compared with a matched Wilcoxon sign rank test implemented via MATLAB's statistics tool box. Doing the comparison at a dataset level allowed us to obtain a discrimination statistic for every metric we used without repeating the computationally expensive analysis procedure hundreds or thousands of times on every Vm trace. With enough individual recordings in each dataset the matched Wilcoxon sign rank test is a reliable measure, which empowered us to efficiently compare all important metrics.

2.2.6 Neuronal Avalanche Analysis

Neuronal avalanches were defined by methods analogous to (Poil et al., 2012), which are used for uninterrupted ongoing signals whereas methods based on event detection (Beggs and Plenz, 2003) require periods of non-activity. A threshold is defined, and an avalanche starts when the signal crosses the threshold from below and ends when the signal crosses the threshold from above. The choice of threshold is a free parameter and we set it to the 25^{th} percentile before conducting the complete analysis. In similar situations (continuous non-zero signals) researchers chose ½ the median (Poil et al., 2012; Larremore et al., 2014). However, ½ the median cannot work for negative signals or signals with high mean but low variance. Before analysis threshold choices between the 15^{th} to 50^{th} percentile

were tested on data from the five cells with the most recordings to see how threshold may affect the number of avalanches. The 25th percentile was in keeping with the existing literature and gave many avalanches compared to alternatives. Having a large number of avalanches is important because it gives the best statistical resolution. An analysis with a choice of threshold that yields fewer avalanches (or changing the threshold for each recording) would be suspect for selecting serendipitous results. After the analysis was conducted eight percentiles between the 15th to 50th were tested and gave similar power-law exponents.

We quantified each neuronal avalanche by its size A and its duration D. The avalanche size is the area between the processed V_m recording and the baseline. The baseline is another free parameter that was set at the second percentile of the processed V_m recording. The second percentile was chosen because its value is more stable than the absolute minimum. The avalanche duration D is the time between threshold crossings.

The lower limit of avalanche duration is defined by the membrane time constant which has been reported to be between 50 and 140 ms for the turtle brain at room temperature (Ulinski, 1990; Larkum et al., 2008). We took a conservative approach by setting the limit at less than half the lower bound on membrane time constant which was significantly less than the lower cut-off from power-law fits. Only avalanches of duration larger than 20 ms were included in the analysis. Thus, we avoided artificially retaining only the events most likely to be power-law distributed.

Following the procedure described above, each processed V_m recording of uninterrupted ongoing activity (i.e., a recording of 2 to 5 minutes duration) yielded 327 ± 148 (mean $\pm$ standard deviation) avalanches. This is insufficient for rigorous statistical fitting on recordings individually (Clauset et al., 2009). Therefore, we grouped avalanches from multiple recordings of ongoing activity of the same cells. Each cell produced between 3 and 19 recordings of ongoing activity (2 to 5 minutes duration each recording), with trials recorded intermittently over a period of 10 to 60 minutes. We grouped recordings based on whether they occurred in the first or second 20-minute period since the beginning of recording from that neuron. Then all the avalanches from the first or second 20-minute period were grouped together with one data object (the group) storing the size, and duration of each avalanche. It is rare for neurons to have recordings in the third 20-minute periods, so this data was not included. Since there was a slow drift in the mean membrane potential over a period of several minutes, we scaled the avalanche sizes from each recording to have the same median as other recordings from the same group. Z-scoring was not useful for accounting for trial to trial variability because it does affect whether a specific time-point is above or below a certain percentile threshold. Therefore, it is not useful for removing variability in avalanche duration. Windowed z-scoring introduces artifacts near action potentials. On average 4 recordings were possible in each 20-minute period. There were 51 neurons with multiple recordings of ongoing activity in the first 20-minutes of experimentation (thus 51 recording groups). Of these, 18 neurons had an additional 20-minute period with more than one recording. This produced a total of 69 groups with 1346 ± 1018 (mean $\pm$ standard deviation) avalanches for each group. Of these 69 groups, 57% had more than 1000 avalanches. The largest number of avalanches was

7495 and the smallest was 313. Only 5 groups had less than 500 avalanches. We report on the 51 groups from the first 20-minute period separately from the 18 groups with recordings from the second 20-minute period of experimentation.

For each group, we evaluated the avalanche size and duration distributions with respect to power laws. To test whether a distribution followed a power law, we applied the rigorous statistical fitting routine described previously (Clauset et al., 2009). We tested three power-law forms: $P(x) \propto x^{-\alpha}$ (with and without truncation) (Deluca and Corral, 2013), as well as a power-law with exponential cut-off $P(x) \propto x^{-\alpha} e^{-x/r}$. We compared these against lognormal and exponential alternative (non-power-law) hypotheses. Distribution parameters were estimated using Maximum Likelihood Estimation (MLE) and the best model out of those fitted to the data was chosen using the Akaike Information Criterion (Bozdogan, 1987). It should be acknowledged that a small power-law region in the truncated form would be suspect for false positives, likewise for a strong exponential cut-off (Deluca and Corral, 2013). Finally, to decide whether a fitted model was plausible, pseudo-random datasets were drawn from a distribution with the estimated parameters and then the fraction which had a lower fit quality (Kolmogorov-Smirnov distance) than the experimental data was calculated. If this fraction, called the comparison quotient q, was greater than 0.10, the best fit model (according to the Akaike Information Criterion) was accepted as the best candidate. Otherwise, the next best model was considered.

We applied several additional steps and strict criteria to control for false positives. One such step was assessing whether the scaling relation was obeyed over the whole avalanche

distribution for each group (not just the portion above the apparent onset of power-law behavior). The scaling relation is another power-law $\langle A \rangle(D) \propto D^{\gamma}$ predicting how the measured size of avalanches increase geometrically with increasing duration (on average). For any data set which has three power-laws, $\langle A \rangle(D) \propto D^{\gamma}$(scaling relation), $P(A) \propto A^{-\tau}$ (size distribution), and $P(D) \propto D^{-\beta}$ (duration distribution), the scaling relation exponent is predicted by the other two exponents by $\gamma \approx \gamma_p = \frac{(\beta-1)}{(\tau-1)}$ (Scarpetta et al., 2018). Note that $\gamma_p = 1$ is a trivial value because it implies $\langle A \rangle(D) \propto D$ and that would suggest individual avalanches were just noise symmetric about a constant value. This would mean that the average avalanche shape is just a flat line at some constant of proportionality, $F\left(\frac{t-t_0}{D}\right) = a$, where $F\left(\frac{t-t_0}{D}\right)$ is a function describing the shape of an avalanche of duration D and t_0 is the beginning of the avalanche and a is a constant.

Standards for consistency with critical point behavior

We applied four standardized criteria to provide a transparent and systematic way to produce a binary classification, either "no inconsistencies with activity near a critical point were detected" or "some inconsistencies with activity near a critical point were detected".

First, a collection of avalanches must be power-law distributed in both its size and duration distributions.

Second, the collection of avalanches must have a power-law scaling relation as determined by $R^2 > 0.95$ (coefficient of determination) for linear least squares regression to a log-log

plot of average size vs durations: $log\ log\left(\langle A\rangle(D)\right) \sim \gamma\ log\ log\ (D)\ + b$. This R^2 represents the best that any linear fit can achieve and must include all the avalanches, not a subset. We denote the scaling exponent (slope from linear regression) from this fit as γ_f.

Third, the scaling relation exponent predicted by theory (denoted as γ_p) must correspond to a trendline on a log-log scatter plot of $\langle A\rangle(D)$ whose R^2 is within 90% of the best-case fitted trendline from the second criterion. Again, the R^2 for the predicted scaling relation is calculated across all avalanches, and not just the subset above the inferred lower cut-off of power-law behavior (which was found for the first criterion). This cross-validates agreement with theory.

Fourth, the fitted scaling relation exponent must be significantly greater than 1: $(\gamma_f - 1) > \sigma_{\gamma_f}$ where σ_{yf} is the standard error. This last requirement eliminates scaling that might be trivial in origin. It is measured after getting the fitted scaling relation exponent for all the data so that a dataset standard deviation can be determined. It is necessary to also check that the set of scaling relation exponents from the power-law fits to all avalanche sets is significantly different from 1 at a dataset level. A scaling relation exponent equal to one suggests a linear relationship between mean-size and duration which is not consistent with criticality in neural systems (Haldeman and Beggs, 2005).

Our four-criterion test cannot measure distance from a critical point nor eliminate all risk of false positives. To complete our analysis, we also look at three additional factors, whether exponent values match exponent values from *other* experiments as expected from

the universality prediction of theory, whether all the exponents within our data set have similar scaling relation predictions, and lastly whether the avalanches within our data set exhibit shape collapse across all the recordings.

Applying shape collapse, quantitative and qualitative analysis.

Shape collapse is a very literal manifestation of scale-invariance (also called "self-similarity")(Sethna et al., 2001; Beggs and Plenz, 2003; Friedman et al., 2012; Pruessner, 2012; Timme et al., 2016). Avalanches of different durations should rise and fall in the same way on average. This average avalanche profile is called a scaling function. The average avalanche profile for avalanches of duration D is predicted to be $A(t, D) = D^{(\gamma-1)}F\left(\frac{t-t_0}{D}\right)$ where $D^{(\gamma-1)}$ is the power-law scaling coefficient which modulates the height of the profile and $F\left(\frac{t-t_0}{D}\right)$ is the universal scaling function itself (normalized in time). Shape collapse analysis provides an independent estimate of the scaling relation exponent γ_{SC}, which is only expected to be accurate at criticality (Sethna et al., 2001; Scarpetta and Candia, 2013; Shaukat and Thivierge, 2016), and a visual test of conformation to an empirical scaling function.

Exponent estimation is very sensitive to the unrelated, intermediate rescaling steps involved in combining the avalanches from multiple recordings into one group. To get an estimate of the scaling relation exponent for each group, γ_{SC}, we average the scaling exponents γ_i found individually for each recording in that group (i denotes the i^{th} recording, SC for "shape collapse").

Naturally, individual avalanche profiles are vectors of variable length D. We must first "rescale in time" to make them vectors of equal length without losing track of what each vector's original duration was. We do that by linearly interpolation with 20 evenly spaced points. So, the j^{th} avalanche profile of the i^{th} recording is denoted as a 20-element vector $\overrightarrow{\Gamma_{ij}}$ (where the top arrow denotes a vector).

Next, the set of all profiles from recording i with the *exact same duration* D, denoted as $\Gamma_{D\,i}$ where bold indicates a set, were averaged and divided by a test scaling factor $D^{(\gamma_i'-1)}$. We define this as $\overrightarrow{\Gamma_{D_i}}(\gamma') = \overrightarrow{\langle\Gamma_{D_i}\rangle}D^{-(\gamma_i'-1)}$. The prime indicates a test rescaling. The average is over all vectors in the set $\Gamma_{D\,i}$. The choice of γ_i was optimized using MATLAB's fminsearch function to minimize the mean relative error between the average over all durations $\langle\overrightarrow{\Gamma_{D_i}}(\gamma')\rangle$ and the set members $\overrightarrow{\Gamma_{D_i}}(\gamma')$ so that for recording i:

$$\gamma_i = \langle \frac{|\overrightarrow{\Gamma_{D_i}}(\gamma') - \langle\overrightarrow{\Gamma_{D_i}}(\gamma')\rangle|}{\langle\overrightarrow{\Gamma_{D_i}}(\gamma')\rangle} \rangle \,. \qquad (2.1)$$

This error minimization and applying the rescaling is the "collapse" in "shape-collapse".

Once we have the γ_i for the avalanches in each individual recording of ongoing activity we compare the average, $\gamma_{SC} = \langle\gamma_i\rangle$, to the predicted and fitted scaling relation exponents for the group of recordings, γ_p and γ_f (statistical comparison tests are described in a previous section). Thus, quantitative analysis of shape collapse was done by comparing γ_{SC}, γ_p, and γ_f for each of the 69 groups individually.

Visual assessment of how well avalanche profiles can be described by one universal scaling function, $F\left(\frac{t-t_0}{D}\right)$ supports the quantitative exponent estimation. This was carried out by averaging all the profiles within specific *duration bins* (regardless of trial or group) and plotting them on top of one another. A very large number of avalanches are needed so we combine avalanches from all 69 groups. However, the resting membrane potential differs from recording to recording and cell to cell. Therefore, avalanche profiles from different recordings are vertically misaligned. To combine avalanches profiles from different recordings we divided all the profiles by a scalar value unique to each recording: the time average over all the collapsed profiles. This produce rescaled and mean-shifted profiles (double prime) $\overrightarrow{\Gamma''_{ij}} = \overrightarrow{\Gamma'_{ij}}/\langle \Gamma'_{ijk} \rangle$ (where $k \in [1,20]$ denotes the interpolated time point). The set of avalanches from each recording were thus aligned, but individual variability was preserved and thus profiles from different recordings could be averaged without introducing artifacts. This set, Γ''_{ij} contained a total of 106,220 shifted and rescaled profiles for the V_m data.

The set of shifted and rescaled profiles falling into a duration bin is denoted Γ''_D. Each duration bin then provides its own estimate of the scaling function $\overrightarrow{\langle \Gamma''_D \rangle} \sim F\left(\frac{t-t_0}{D}\right)$. For each bin, D was defined as the average duration of all constituent profiles. If less than 700 avalanches had a particular duration, we included the next longest duration iteratively until we met or exceeded 700 avalanches. This only applied to long durations. The choice of 700 was made because it allowed us smooth averaging and without excessively wide duration bin widths.

We also assessed the mean curvature of avalanche profiles from the rescaled profile for a particular duration $\overrightarrow{\langle \Gamma_D'' \rangle}$. This allows us to plot how curvature depends on duration. Mean curvature $\langle \kappa \rangle$ is defined like so (k still denotes time points):

$$\langle \kappa \rangle(D) = \left\langle \frac{\ddot{F}\left(\frac{t-t_0}{D}\right)}{\left(1+\dot{F}\left(\frac{t-t_0}{D}\right)\right)^{\frac{3}{2}}} \right\rangle \sim \left\langle \frac{\langle \Gamma_D'' \rangle_{(k+1)} - 2\langle \Gamma_D'' \rangle_k + \langle \Gamma_D'' \rangle_{(k-1)}}{\left(1+\langle \Gamma_D'' \rangle_{(k+1)} - \langle \Gamma_D'' \rangle_k\right)^{\frac{3}{2}}} \right\rangle. \qquad (2.2)$$

2.2.7 Model Simulations

We simulated a model network consisting of $N = 10^4$ binary probabilistic model neurons. The model neurons form a directed random network (Erdős–Rényi random graph), where the probability that neuron j connects to neuron i is c. In a network of N neurons, this results in a mean in-degree and out-degree of cN. We tested nine not quite evenly distributed values of connection probabilities $c \in [0.5,1,3,5,7.5,10,15,20,25] \times 10^{-2}$. As discussed in (Kinouchi and Copelli, 2006; Larremore et al., 2011a, 2014) the impact of connectivity on network dynamics is non-linear, so we take a finer look at smaller connection probabilities, while maintaining thorough coverage of intermediate connection probabilities.

The strength of the connection from neuron j to neuron i is quantified in terms of the network adjacency or weight matrix W with the fortune of having a simple and intuitive meaning. For each existing connection from neuron j to neuron i, W_{ij} is the direct change in the probability that neuron i will fire at the next timestep if neuron j spikes in the current time step.

The dynamics of this network is well-characterized by the largest eigenvalue λ of the network weight matrix W, with criticality occurring at $\lambda = 1$ (Kinouchi and Copelli, 2006; Larremore et al., 2011a, 2011b, 2012, 2014). The physical interpretation of λ is a "branching parameter"(Haldeman and Beggs, 2005) that governs expected number of spikes immediately caused by the firing of one neuron. If $\lambda = 1$ then one spike causes one other spike on average, while if $\lambda > 1$ one spike causes more than one on average and vice versa.

We tested five different values of largest eigenvalue at, near and far from criticality $\lambda \in [0.9,0.95,1,1.015,1.03]$. A fraction χ of the neurons are designated as inhibitory. This is done by multiplying all outgoing connections of an inhibitory neuron by -1. We tested nine different values of the fraction of inhibitory neurons in the range from 0 to 0.25, thus including the value 0.2, corresponding to the fraction of inhibitory neurons in the mammalian cortex (Meinecke and Peters, 1987). The magnitudes of non-zero weights are independently drawn from a distribution of positive numbers with mean η, where the distribution is uniform on $[0,2\eta]$, and η is given by $\eta = \lambda/\big(cN(1-2\chi)\big)$. The maximum eigenvalue is then fine-tuned by dividing W by the current maximum eigenvalue and set to the exactly desired value $W = \lambda W'/\lambda'$ where W' and λ' are the matrices and eigenvalues before correction.

The binary state $S_i(t)$ of neuron i at time t denotes whether the model neuron spikes ($S_i(t) = 1$) or does not spike ($S_i(t) = 0$) at time t. At each time step, the states of all neurons are updated synchronously according to the following update rule:

$$S_i(t) = \Theta\left(\sum_j^N W_{ij} S_j(t-1) - \xi_i(t)\right), \qquad (2.3)$$

where $\xi_i(t)$ is a random number on [0 1] drawn from a uniform distribution, and Θ is the Heaviside step function. In addition to this update rule, a refractory period of one time-step (translated to approximately 2 ms) was imposed for certain parameter conditions. A simulation begins with initiating the activity of one randomly-chosen excitatory neuron and continuing the simulation until overall network activity had ceased. The process was then repeated.

From the simulated binary states of 10^4 model neurons, we extracted three measures of simulated activity. First, the network activity $F(t) = \sum_{i=1}^{N} S_i(t)/N$ is the fraction of neurons spiking at time t. Second, the input to model neuron i at time t is $P_i(t) = \sum_j^N W_{ij} S_j(t-1)$, which is almost always positive for our parameters. Note that $P_i'(t) = P_i(t) \times \Theta\left(P_i(t)\right)$ directly represents the probability for the neuron to spike at time t. Third, we constructed a proxy for the V_m signal, $\Phi_i(t) = (\alpha_h * P_i)(t)$, by convolving the input $P_i(t)$ with an alpha function: $\alpha_h(t) = \frac{t}{h_m} exp\ (1 - \frac{t}{h_m})$ with $h_m = 2$ time steps (assumed to be about 4 ms).

A total of 405 different parameter combinations (connection density, inhibition, maximum eigenvalue) were simulated. Each combination was simulated 10 times. Based on the connection probability c and the fraction of inhibition χ, we distinguish four regions in parameter space classified according to the behavior of the critical model, i.e., $\lambda = 1$.

The first region is the "positive weights" region. Without inhibition activity increases or dies out in accordance with the branching parameter. This region is defined by $\chi = 0$. With moderate inhibition and dense connectivity there is a region of parameter space we call "quiet"; activity lasts only slightly longer than in a system with no inhibition. This region is defined by the ex-post-facto boundaries $c \geq e^{11\chi}/25$ and $\chi > 0$. Further increasing inhibition relative to connection density produces a behavior like "up and down" states (or "telegraph noise") (Sachdev et al., 2004; Millman et al., 2010). We call this the "switching" regime because network activity switches between a low mean and a high mean. This region is defined by $< e^{11\chi}/25$, and $c \geq (10e^{12\chi} - 13)/100$ and $\chi > 0$. When inhibition is high relative to connection density the system enters the "ceaseless" region where stimulating one neuron causes activity that effectively never dies out. An especially attractive feature of this model is that the "ceaseless" and "switching" regimes exhibit sustained self-generated activity. This provides a way to model spontaneous neural activity without externally imposed firing patterns.

Refractoriness was studied in the network without inhibition and it was found that dynamic range was inversely proportional to refractory period (Larremore et al., 2011a) but the branching parameter (criticality) displayed no dependence on refractory period (Kinouchi and Copelli, 2006). In the literature which featured inhibition and introduced ceaselessness no refractory was used (Larremore et al., 2014). However, we found that for some networks in the switching regime the maximum eigenvalue was a better predictor of the empirical branching ratio if the refractory period was one timestep. Because this relationship is central to our understanding of criticality in this model, we ran an initial

testing cycle before each simulation begins to decide whether to set the refractory period to one timestep or zero. Doing so ensures the network displays critical-like phenomena in all regimes (the maximum eigenvalue of connectivity) but also ensures the model adheres to the practices of the literature.

We performed avalanche analysis on each of the simulated signals using the methods described above for membrane potential recordings. If the network is in the switching regime, we only perform analysis on the periods when the network is in the mode (high or low mean) in which it spends the majority of its time. As before, the 25th percentile defined the avalanche threshold. If the signal had negative values, as in the case of single neuron Vm proxies in networks with inhibition, the signal was shifted by subtracting the 2nd percentile. To obtain good statistics, we continued stimulating and extracting avalanches until a simulation either reached 10^4 avalanches, or 5×10^3 avalanches and a very large file size or a very long computational time. This ensured there were between two and ten thousand avalanches per trial.

2.2.8 Data and Software Accessibility

All raw data is available at https://github.com/jojker/continuous_signal_avalanche_analysis and the software developed for this analysis is available upon request to the author: James Kenneth Johnson.

2.3 Results

Single-neuron membrane potential (V_m) fluctuations are thought to be dominated by synaptic inputs from multitudes of presynaptic neurons (Stepanyants et al., 2002; Brunel et al., 2014; Petersen, 2017). Since the way neurons integrate their diverse inputs is central to information processing in the brain, it is important that neuroscience gain a thorough understanding of the relationship between subthreshold V_m fluctuations and population activity. A basic step is to compare statistical analyses, especially analyses where a meaningful relationship is expected. We asked whether an avalanche analysis on V_m fluctuations would reveal the same signatures of scale-freeness and critical network dynamics found in measures of population activity (Figure 2.1) (Friedman et al., 2012; Shew et al., 2015; Marshall et al., 2016). To address this comparison across organizational levels, we recorded V_m fluctuations from 51 pyramidal neurons in visual cortex of 14 turtles and assessed evidence for critical network dynamics from these recordings.

In a model investigation we corroborated results evaluated the conditions needed to enable inferring dynamical network properties from the inputs to single neurons. Finally, we extended the analysis to other commonly recorded time series of neural activity for comparison with the information content of V_m fluctuations about the dynamical network properties.

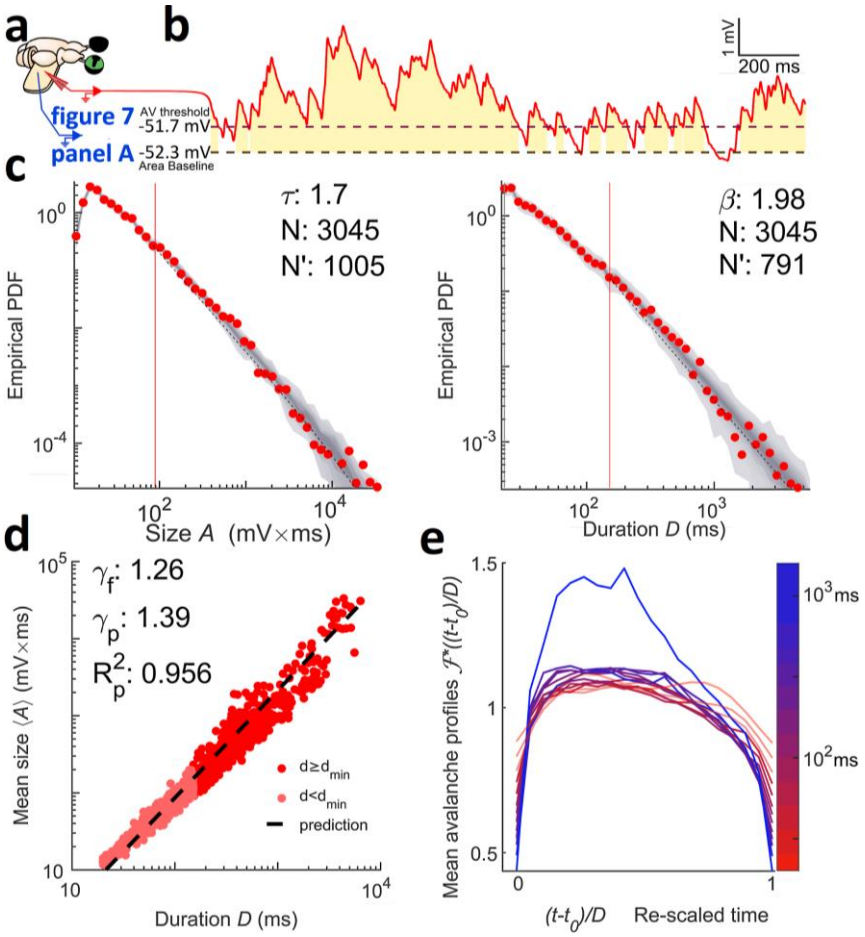

Figure 2.2 | Membrane potential fluctuations reveal signatures of critical point dynamics. Panel a shows the whole-brain eye attached joint V_m and LFP recording preparation. **Panel b** shows that the membrane potential (red) is thresholded at the 25th percentile (a dashed line). Avalanches are defined by excursions above this threshold. The gold region represents the size of the avalanche, which is the area between the signal and its 2nd percentile (a dashed line). The duration of the avalanche is the duration of the excursion. **c** shows the size (left) and duration (right) distributions of V_m inferred avalanches when data is combined from seven recordings from the same neuron falling in the same 20-minute period. The comparison quotients (q) are both above 0.10 (0.878 and 0.874 respectively), indicating that the size and duration distributions were better fits to power-laws at the given cut-off than 87% of power-laws produced by a random number generator with the same parameters (shown as a grey density cloud). N' indicates the number of avalanches above the lower cut-off of the fit (red vertical line) and N indicates the total number of avalanches. Size duration exponent denoted with τ while β is used for duration. **d** shows the scaling relation which is a function relating average avalanche size to each given duration. The predicted exponent (γ_p) successfully explains 95.6% of the variance of a log-log representation of the data. A linear least squares regression could explain 96.7% and gives the fitted exponent (γ_f). Therefore, γ_p comes within 1.2% of the best linear explanation despite a 10% difference in exponent values. **e** shows shape collapse. Each line represents the average time-course of an avalanches of a given duration. The color indicates the duration according to the scale bar. Durations below 50 ms (the lower bound on turtle pyramidal time-constants) are made translucent and slightly thickened. This shape collapse represents the global collapse across all recordings in all cells. This confirms that a universal scaling function,

$F\left(\frac{t-t_0}{D}\right)$, is present. For the seven recordings in the group represented in panels C & D, the mean scaling relation exponent derived from shape collapse was $\gamma_{SC} = 1.23$ a disagreement of 2.2% relative to γ_f.

2.3.1 Membrane Potential Fluctuations Reveal Signatures of Critical Point Dynamics

We obtained whole-cell recordings from pyramidal neurons in the visual cortex of the turtle ex-vivo eye-attached whole-brain preparation (Figure 2.2a). Recorded V_m fluctuations taken in the dark (no visual stimulation) were interpreted as ongoing activity. We analyzed the recorded ongoing V_m fluctuations employing the concept of "neuronal avalanches" (Beggs and Plenz, 2003; Poil et al., 2012; Shew et al., 2015), which are positive fluctuations of network activity. For continuous time-series such as the V_m recording, one selects a threshold and a baseline. We defined a neuronal avalanche based on the positive threshold crossing followed by a negative threshold crossing of the V_m time series (Poil et al., 2012; Hartley et al., 2014; Larremore et al., 2014; Karimipanah et al., 2017a). We quantified each neuronal avalanche by (i) its size A, i.e., the area between the curve and the baseline, and (ii) its duration *D*, i.e., the time between threshold crossings (Figure 2.2b).

To quantify the statistics of avalanche properties, we applied concepts and notations from the field of "critical phenomena" in statistical physics (Nishimori and Ortiz, 2011; Pruessner, 2012). Because the critical point is such a small target for any naturally occurring self-organization (Pruessner, 2012; Hesse and Gross, 2014; Cocchi et al., 2017) and there is considerable risk of false positives (Taylor et al., 2013; Hartley et al., 2014; Touboul and Destexhe, 2017; Priesemann and Shriki, 2018), asserting criticality in a *new*

system or with a *new* tool requires extraordinary evidence. Since this is a new tool, we created four criteria and set quantifiable standards for concluding a system is consistent with criticality based on avalanche power-laws and we completed this exhaustive battery of tests with shape collapse, a geometrical analysis of self-similarity in the avalanche profiles (see Methods: Experimental Design and Statistical Analysis).

In brief, we found that both the size and duration distributions of the fluctuations treated as avalanches were consistent with power laws (Figure 2.2c), $P(A) \propto A^{-\tau}$ and $P(D) \propto D^{-\beta}$ matching widely reported exponents (Beggs and Plenz, 2003; Priesemann et al., 2009; Hahn et al., 2010; Klaus et al., 2011; Friedman et al., 2012; Shriki et al., 2013; Priesemann, 2014; Arviv et al., 2015; Shew et al., 2015; Karimipanah et al., 2017b, 2017a), obeyed the scaling relation (Figure 2.2d), and exhibited shape collapse over an expansive set of durations, (Figure 2.2e).

Specifically, of the 51 recording groups featuring data from the first 20-minute period of recording from one cell, 98% had power laws in both size and duration distributions. The exponent values for the size distribution were $\tau = 1.91 \pm 0.38$ (median $\pm$ standard deviation). Exponent values for the duration distribution were $\beta = 2.06 \pm 0.48$. Of the 51 neurons with a recording group from the first 20-minutes, 18 had an additional 20-minute period spanning multiple recordings. All of these 18 groups had power-laws in both size and duration, the exponent values for the size distribution were $\tau = 1.87 \pm 0.29$ and the exponent values for the duration distribution were $\beta = 2.21 \pm 0.39$.

It is also important to confirm that power-law behavior extends across several orders of magnitude of avalanche durations. We typically demonstrate a power-law distribution over 2.45 ± 0.39 orders of magnitude of duration. For the scaling relation we find a larger span with 2.62 ± 0.23 orders of magnitude across our whole avalanche duration range.

Another statistic crucial to signatures of criticality measures the relationship *between* the power-laws describing size and duration of avalanches (Sethna et al., 2001; Beggs and Timme, 2012; Friedman et al., 2012). If the average avalanche size also scales with duration according to $\langle A \rangle(D) \propto D^{\gamma}$, then the exponent γ is not independent, but rather depends on the exponents τ and β according to $\gamma = (\beta - 1)/(\tau - 1)$ irrespective of criticality (Scarpetta et al., 2018). For critical systems this condition is enforced because avalanche profiles follows the same shape for all durations which means that this prediction is believed to be more precise than for non-critical systems and the exact values are important (Sethna et al., 2001; Nishimori and Ortiz, 2011). We found that average avalanche size scaled with duration $\langle A \rangle(D) \sim D^{\gamma}$ according to a power law and that the observed values of τ and β provided a good prediction $\gamma = (\beta - 1)/(\tau - 1)$ of the fitted γ (Figure 2.2d).

Specifically, of the 51 recording groups from the first 20-minute period, the fitted scaling relation exponents were $\gamma_f = 1.19 \pm 0.05$, and the predicted scaling relation exponents were $\gamma_p = 1.17 \pm 0.35$. For the additional second 20-minute period (18 groups/neurons), the fitted scaling relation exponents were $\gamma_f = 1.21 \pm 0.05$, and the predicted scaling relation exponents were $\gamma_p = 1.28 \pm 0.21$.

To affect a more convincing analysis, we defined four stringent criteria that must be independently satisfied before any set of avalanches can be deemed consistent with network dynamics near a critical point (see Methods: Experimental Design and Statistical Analysis). Overall, of the 69 groups of recordings (which includes 18 out of 51 cells twice), 98.6% had power-laws in both the size and duration distributions of avalanches and 92.8% had scaling relations which were well fit by power-laws ($R^2 > 0.95$). All were deemed non-trivial by the test $(\gamma_f - 1) > \sigma_{\gamma_f}$ where σ_{γ_f} is the dataset standard error; $\sigma_{\gamma_f} = 0.051$. The smallest value was $\gamma_f = 1.094$. The fourth constraint, that the R^2 of the predicted scaling relation was within 10% of the best fit scaling relation, was satisfied 85.6% of the time. Together, this set of criteria cannot measure distance from a critical point nor eliminate false positives. However, the take away is that 81% of all recording groups examined were judged to be consistent with network activity near a critical point.

Separating out results: 76% of the 51 recording groups from the first 20-minute period, and 94% of the recording groups from the second 20-minute period were judged consistent with criticality. The general pattern is that the first 20-minute period and the second are both consistent with criticality, but the second group meets our criteria much more frequently. This could be an effect related to the length of time we are able to maintain a patch, or it could be that a better patching results in both longer stable recording ability and better inference of dynamical network properties.

To further discount the possibility of false positives we investigated whether the avalanches within our data set exhibited “shape collapse” (Figure 2.2e). The scaling

relation is a consequence of self-similarity (Sethna et al., 2001; Papanikolaou et al., 2011; Friedman et al., 2012; Marshall et al., 2016; Shaukat and Thivierge, 2016; Cocchi et al., 2017). In other words, avalanches all have the same "hump shape" no matter how long they last, this shape is called the scaling-function or avalanche profile. The shape collapse also provides an independent estimate of the scaling relation exponent γ, if the estimated exponent, γ_{SC}, matches the fitted exponent, γ_f, it is considered strong evidence of critical point behavior. For critical systems, the average avalanche profile of an avalanche of duration D is given as $A(t, D) = D^{(\gamma-1)}F\left(\frac{t-t_0}{D}\right)$. Where $D^{(\gamma-1)}$ is a coefficient governing the scaling of height with duration, and $F\left(\frac{t-t_0}{D}\right)$ is the scaling-function which describes the universal shape of an avalanche at any duration. The similarity of avalanche profiles of different durations is qualitatively judged (Sethna et al., 2001; Beggs and Plenz, 2003; Friedman et al., 2012; Pruessner, 2012; Timme et al., 2016) by plotting empirically estimated scaling functions for several durations on top of one another after they have been rescaled as part of the process of estimating γ_{SC}.

We obtained shape collapse across more than one order of magnitude (between about 50 ms to 700 ms) of avalanche durations. Below 50 ms distinct peaks arose. Above 700 ms the profile height grew faster than the power-law scaling that worked for shorter duration avalanches, this is observed as an apparent outlier in Figure 2.2e. This likely marks point where avalanches become so long and so large that they begin to weakly activate the non-linear action potential mechanism of the neuron. When comparing to plausible alternatives to V_m in later sections, we included analysis of mean curvature and avalanche profile peak height along with visual inspection of shape collapse quality (Figure 2.2e). The shape

collapse plots begin with short avalanches (20 ms) that are below the median lower cut-off for power-law behavior (which was 256 ms) but are well predicted by the scaling relation.

The exponents estimated from the shape collapse were a good match for both the predicted and fitted scaling relation exponents. The groups of recordings from the first 20 minutes yielded $\gamma_{SC} = 1.1868 \pm 0.042$. The average matched absolute percent error was 1.3% with respect to γ_f. A matched signed rank difference of median test revealed that γ_f was not significantly different from γ_{SC}, simple difference effect size $r_{SDF} = 0.089$, p-value $p = 0.063$ ($p < 0.05$ indicates that they are different).

This stage of the analysis showed that, when fluctuations of V_m are treated like neuronal avalanches, they are consistent with criticality by the standards of power-laws governing size and duration. We also showed that V_m avalanches exhibit geometrical self-similarity across more than one order of magnitude. These factors showed that the cortical circuits driving fluctuations of membrane potential are consistent with activity near a critical point according to standards of self-similarity. In our next investigation we compared to population data from microelectrode arrays and other results from literature to test whether V_m fluctuations are consistent with the universality requirement of behavior near critical points, and whether they can be used to measure dynamical network properties.

2.3.2 Membrane Potential Fluctuations are Consistent with Avalanches from Previously Obtained Microelectrode Array LFP Recordings

Importantly, we sought to interpret our results from the analysis of single-neuron V_m fluctuations in the context of the more commonly used analysis of multi-unit spiking activity (Friedman et al., 2012; Shew et al., 2015; Marshall et al., 2016; Karimipanah et al., 2017a) or multi-site local field potential (LFP) event detection from microelectrode array (MEA) data (also known as "multielectrode array") (Beggs and Plenz, 2003; Shew et al., 2015).

In a previous study, avalanche analysis was performed on LFP multi-site MEA recordings from the visual cortex of a different set of 13 ex-vivo eye-attached whole-brain preparations in turtle (Shew et al., 2015). Avalanches were inferred from the steady state (after on response transients but before off response transients) of responses to visual presentation of naturalistic movies as opposed to the resting state activity between presentations (which is where the V_m data come from). Avalanche size and duration distributions followed power laws.

The median exponents were $\tau = 1.94 \pm 0.27$ for the avalanche size distributions and $\beta = 2.14 \pm 0.32$ for the avalanche duration distributions (Figure 2.3a). A scaling relation existed with average exponent $\gamma_f = 1.20 \pm 0.06$ fitted to the data and $\gamma_p = 1.19 \pm 0.07$ from the average of the predicted scaling based on theory. The scaling power-law extended over 1-2 orders of magnitude. Critical branching was more firmly established in Shew et al., 2015 by analyzing the branching ratio. The branching ratio is the average ratio of events (i.e. spikes) from one moment in time to the next, but only during identified avalanches. A critical branching network has a branching ratio of one, but empirically estimating it

requires discrete events and an assiduous choice of time-binning for analysis. Shew et al. found that a branching ratio near one that was robust to reasonable choices of time bin and varied with choice of time-bin in expectation with critical branching. We are not aware of methods for estimating a branching ratio in continuous signals like membrane potential.

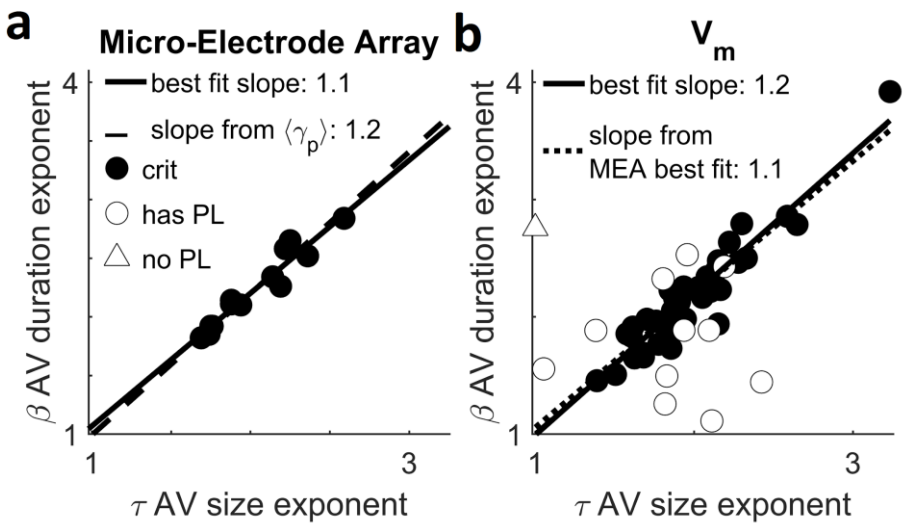

Figure 2.3 | Membrane potential fluctuations are consistent with avalanches from previously obtained microelectrode array data. A plot of the exponents governing power-law scaling of avalanche duration vs the exponents governing avalanche size. Circles indicate data which was best fit to a power-law in both its size and duration. Triangle indicates otherwise (the MLE estimation of a would-be power-law fit, the "scaling index", is plotted in that case (Jeżewski, 2004)). Filled circles indicate data that meet all four standardized criteria for judging data to be consistent with criticality. **a** is a reproduction from (Shew et al., 2015). It shows the results of avalanche analysis on microelectrode array data collected during the steady state of stimulus presentation in an otherwise identical experimental preparation. The exponent values appear to covary to maintain a stable value of the scaling relation $\gamma_p = \frac{\beta-1}{\tau-1}$. The correlation between β and τ was high (see Results: The Predicted Scaling Relation Exponent is More Stable than Avalanche Size or Duration Exponents). **b** shows the results of avalanche analysis performed on fluctuations in subthreshold membrane potential. We found power-laws with closely matching exponents and the same scaling relation with the similar level of stability. The correlation between β and τ was high (see Results: The Predicted Scaling Relation Exponent is More Stable than Avalanche Size or Duration Exponents).

The set of avalanche size, duration, and scaling relation exponents obtained from membrane potential fluctuations (Figure 2.3b) were not distinguishable from the MEA obtained set. The fitted scaling relation exponent γ_f had the least variability of all three

kinds of exponents so it is the most likely to show a difference. Thus, if a difference is not significant it suggests universality more strongly than for the avalanche size τ or duration β distribution exponents.

When we limited our analysis to the first twenty-minute period which contained multiple recordings (51 cells), neither the fitted scaling relation exponent, nor the predicted scaling relation exponent were significantly different from the MEA results. The Wilcoxon rank-sum difference of medians test against the MEA data yielded ($r_{SDF} = 0.164,\ p = 0.37$), and ($r_{SDF} = 0.08, p = 0.67$) respectively. The median exponent values for the size and duration distributions were not significantly different from the median of the MEA data ($r_{SDF} = 0.164, p = 0.37$) and ($r_{SDF} = 204,\ p = 0.265$) respectively.

These results establish V_m fluctuations as an informative gauge of high-dimensional information, while also demonstrating that the power-law characteristics are universal properties of the brain, by showing a close match between data at different scales and under different conditions. Further underscoring universality, our results are also similar to the critical exponents measured from other animals such as the $\tau = 1.8$ result from in-vivo anesthetized cats (Hahn et al., 2010), though an exhaustive literature search was not conducted, others have conducted incomplete surveys (Ribeiro et al., 2010; Priesemann, 2014).

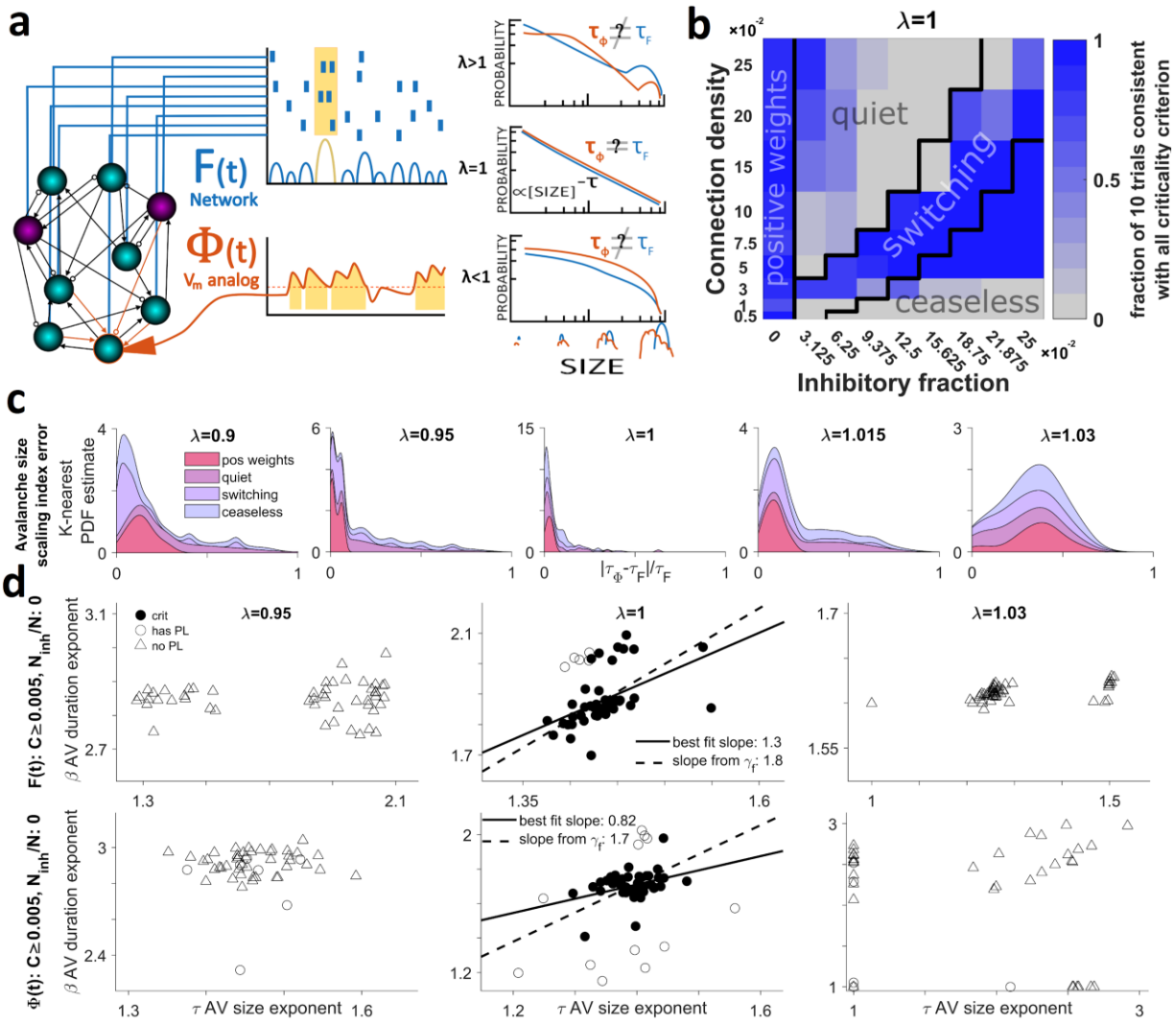

Figure 2.4 | The single-neuron estimate of network dynamics is optimized at the network critical point. a illustrates model network consists of 10^4 excitatory (cyan) and inhibitory (magenta) model neurons with sparse connectivity (line tips: arrows = excitation; circles = inhibition). The simulated model activity (raster plot) is reresented in terms of the single-neuron spiking (raster plot) and the active fraction of the network $F(t) = S(t)/N$ where population spiking is $S(t)$. Concurrently, the smoothed inputs (orange) to a single neuron represents the Vm proxy, $\Phi_i(t)$. The threshold (dashed line) crossings of $\Phi_i(t)$ define avalanches (see Methods: Experimental Design and Statistical Analysis). Avalanches of $F(t)$ and $\Phi_i(t)$ are analyzed in terms of their size (shown) and duration (not shown) distributions and their corresponding exponents, τ. Avalanche statistics depend on several network parameters including the critical branching tuning parameter λ. **b** shows how the inclusion of inhibition affects the network behavior. The black lines mark the boundaries of arbitrarily defined parameter regions roughly corresponding to distinct kinds of behavior. The shade of blue indicates what fraction of ten trials at each point met all four of our standardized criteria for consistency with expectations of critical branching behavior. **c** is a stacked area chart showing the probability density distribution of size exponent error (between $F(t)$ and $\Phi_i(t)$) for different λ and dynamical regimes. The vertical thickness of each color band shows the probability density for that subset of the data while the outer envelope shows the over-all probability density. Probability

density is estimated with a normal kernel smoothing function. In this panel we can see that power-law scaling is most similar at criticality despite variability dependent on the parameter regime. **d** shows a complete summary of the tests for criticality when applied to $F(t)$ (top row) and $\Phi_i(t)$ (bottom row). From this we can confirm that the system is consistent with criticality when there is no inhibition. The subsampling method $\Phi_i(t)$ demonstrates consistency with criticality but displays a wider dispersion of exponent estimates. For experimental V_m and MEA data there was a large correlation between β and τ showing that the scaling relation (which predicts the slope of the trendline) is much more stable than exponent values. This is not the case for the model where for $F(t)$ the correlation is low (see Results: The Predicted Scaling Relation Exponent is More Stable than Avalanche Size or Duration Exponents).

2.3.3 The Single-Neuron Estimate of Network Dynamics is Optimized at the Network Critical Point

To gain a deeper insight into the relation between single-neuron input and network activity, we investigated a model network of probabilistic integrate and fire model neurons (Kinouchi and Copelli, 2006; Larremore et al., 2011a, 2011b, 2012, 2014, Karimipanah et al., 2017a, 2017b). This model network contains fundamental features of cortical populations, such as low connectivity, inhibition, and spiking, while being sufficiently tractable for mathematical analysis (see Methods: Model Simulations).

In brief, the model network consists of $N = 10^4$ binary probabilistic model neurons (Figure 2.4a). The connection probability c results in a mean in-degree and out-degree of cN. The connection strength from neuron j to neuron i is quantified in terms of the network adjacency matrix W. Each connection strength W_{ij} is drawn from a distribution of (initially) positive numbers with mean η, where the distribution is uniform on $[0,2\eta]$. A fraction χ of the neurons are designated as inhibitory, i.e., their outgoing connections are made negative. The binary state $S_i(t)$ of neuron i is updated according to $S_i(t) =$

$\Theta\left(\sum_j^N W_{ij}S_j(t-1) - \xi_i(t)\right)$, where $\xi_i(t)$ is a random number between 0 and 1 drawn from a uniform distribution, and Θ is the Heaviside step function.

The largest eigenvalue $\lambda = \eta cN(1-2\chi)$ of the network adjacency matrix W, characterizes the network dynamics, with critical network dynamics occurring at $\lambda = 1$. This tuning parameter λ controls the degree to which spike propagation "branches": $\lambda = 1$ means that one spike creates one other spike on average, $\lambda > 1$ implies that one spike creates more than one other spike while $\lambda < 1$ means that one spike creates less than one other spike (Haldeman and Beggs, 2005; Kinouchi and Copelli, 2006; Levina et al., 2007; Larremore et al., 2011b, 2012; Kello, 2013).

The input to model neuron i, is $P_i(t) = \sum_j^N W_{ij}S_j(t-1)$ and provides the link between network activity and single-neuron activity. From this we can derive a simple mathematical result characterizing how estimation of network properties is optimized at criticality.

If we let $K_i(t-1)$ denote the number of active neurons *in the presynaptic population* of neuron i, then we can rewrite the input to a model neuron as a sum of independent and identically distributed random variables drawn from the non-zero entries of W: $P_i(t) = \sum_k^{K_i(t-1)} W_{ij_k}$. After implementing inhibition by inverting some elements of W the distribution of weights is not uniform but piecewise uniform. Weights are drawn uniformly from the interval $[-2\eta, 0]$ with probability χ and from the interval $[0,2\eta]$ with probability $1-\chi$. The mean of the nonzero entries of W are denoted with a prime so that the mean is $\langle W'_{ij}\rangle = \eta(1-2\chi)$ and the standard deviation is $\sqrt{\langle W_{ij}^2\rangle - \langle W_{ij}\rangle^2} = \eta\sqrt{\left(1-12(\chi^2-\chi)\right)/3}$.

Now we can find the mean behavior of the input integration function as it relates to the presynaptic population:

$$\langle P_i(t)\rangle \pm \sigma_{P_i(t)} = \eta(1-2\chi)K_i(t-1) \pm \eta\sqrt{\big(1-12(\chi^2-\chi)\big)K_i(t-1)/3}. \qquad (2.4)$$

We learn three things by examining the mean behavior of the input integration function. First, the mean grows as $O(K_i)$ but the standard deviation grows as the root $O\big(\sqrt{K_i}\big)$, so the function becomes a more precise estimator of network activity with increasing activity in the presynaptic population (increasing K_i). Second, the input integration function $P_i(t)$, is rarely negative. At the parameter combination $c = 0.005$ and $\chi = 0.25$ (which has the largest variance relative to the mean) the mean becomes more than one standard deviation larger than zero when $K_i > 5$. Third, and most importantly, the input integration function is an averaging operator and the tuning parameter λ biases that averaging operation. To show this we only need two observations: the instantaneous firing rate averaged over the presynaptic population is the number of active neurons divided by the expected total number of presynaptic neurons, $\omega_i(t) = K_i(t)/cN$. Next, we rearrange the definition of lambda to get $\lambda/cN = \eta(1-2\chi)$. Substituting these two observations into the mean behavior of our input integration function we get the key mathematical result:

$$P_i(t) = \sum_j^N W_{ij}S_j(t-1) \sim \lambda\omega_i(t-1). \qquad (2.5)$$

Note that $P_i'(t) = P_i(t) \times \Theta\big(P_i(t)\big)$ directly represents the probability for the neuron to spike at time t.

These results demonstrate that the inputs to a neuron P_i, and the instantaneous firing rate of that neuron are the result of an averaging operator acting on the presynaptic population, which is a subsample of the network. Furthermore, the tuning parameter λ not only modulates the relationship of single neuron firing to downstream events (also known as branching), but also governs how the input to a neuron relates to the presynaptic population. It biases the averaging operator to either amplify firing rate ($\lambda > 1$) or dampen it ($\lambda < 1$). Therefore, our model implements both critical branching and the inverse of the critical branching condition, a *critical coarse-graining* condition. The model is a network of subsampling operators who only capture whole-system statistics when $\lambda = 1$ and the operators reflect an unbiased stochastic estimate of mean firing rate among the subsample (the presynaptic population). This averaging operation may exist in many kinds of networks, including those with structure and those that are not critical branching networks, so this result helps establish plausible generalizability.

To further evaluate the relation between single-neuron input and network activity under different conditions, we simulated the described network of 10^4 model neurons for a total of 405 different parameter combinations, including connection probability, inhibition, and maximum eigenvalue (Figure 2.4a), each parameter combination was repeated ten times. We then compared the avalanche analysis results of simulated network activity $F(t) = (\frac{1}{N}) \sum_{i=1}^{N} S_i(t)$ and the input to a single neuron (the input integration function). However, $P_i(t) = \sum_j^N W_{ij} S_j(t)$ is the probability that neuron i will fire at time t, also known as the instantaneous firing rate of neuron i.

Membrane potential is not a direct representation of firing rate, but rather the firing rate is related to synaptic input through the F-I curve which is non-linearly related to membrane potential. This non-linearity could destroy the correspondence between the simulated single neuron signal and network activity. In order to better facilitate comparison of the simulated input integration function with the experimentally recorded membrane potential, we constructed a proxy for the subthreshold membrane potential, $\Phi_i(t)$, of a model neuron by convolving the simulated input $P_i(t)$ with an alpha function (see Methods: Model Simulations).

The parameter space has four distinct patterns of critical network behavior (Figure 2.4b). Qualitatively, these were reflected in the network activity. As the presence of these paradoxical behaviors may indicate the presence of second phase-transition tuned by the balance of excitation to inhibition (Shew et al., 2011; Poil et al., 2012; Kello, 2013; Hesse and Gross, 2014; Larremore et al., 2014; Scarpetta et al., 2018) several key results differ strongly and thus are reported separately for these regions of parameter space.

These regions are defined in terms of the connection density and inhibition and shown in Figure 2.4b. First is the "positive weights" region, there is no inhibition ($\chi = 0$) and the network is a standard critical branching network. The second region, "quiet", has a small increase in the fraction of inhibitory neurons. Activity lasts slightly longer than for the classically critical network. The third region is called the "switching" regime because network activity switches between a low mean and a high mean (like "up and down states" (Destexhe et al., 2003; Millman et al., 2010; Larremore et al., 2014; Scarpetta et al., 2018)).

This occurred in the middle portion of the values of connectivity and inhibition. Lastly, we have the “ceaseless” region, with a large fraction of inhibition, relative to connection density, activity never dies out. This region is defined by $c < (10e^{12\chi} - 13)/100$ and $\chi > 0$. Three of these regimes are displayed in Figure 2.5a, the “quiet” region is mostly redundant to the "positive weights" region. The “ceaseless” and “switching” regimes exhibit sustained self-generated activity and is included with the intention to model ongoing spontaneous activity dynamics without contamination by externally imposed firing patterns (Mao et al., 2001).

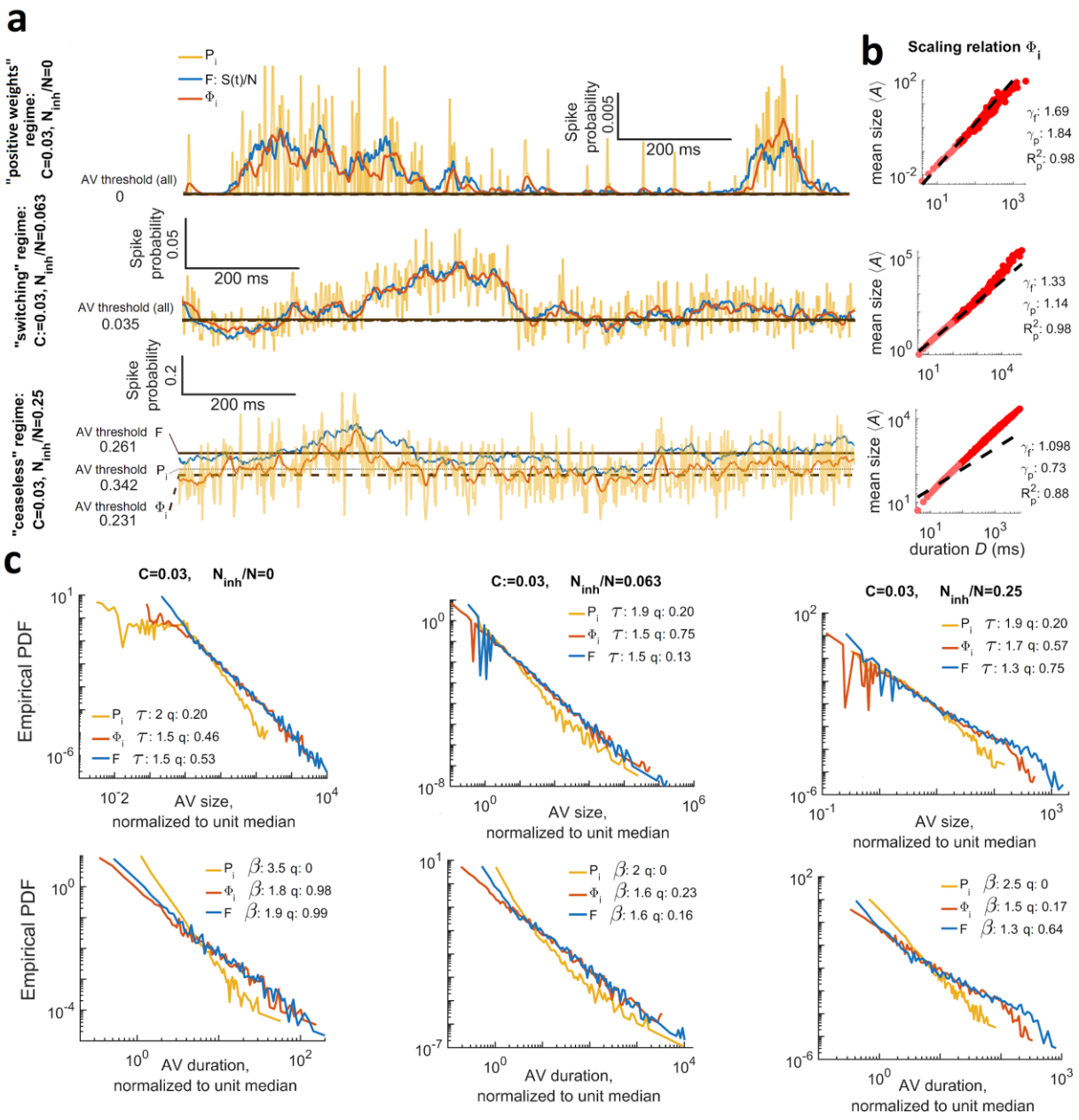

Figure 2.5 | Inputs to a neuron stochastically estimate firing of its presynaptic pool in this critical branching model. a shows differences in model activity dynamics with parameter regions (constant connectivity, $\lambda = 1$, but inhibition, χ varies). Each plot shows the active fraction of the network $F(t)$ in blue, the instantaneous firing rate of node, $P_i(t)$, is in gold and the V_m proxy for the same node, $\Phi_i(t)$, is in orange. The node is randomly selected from the nodes with degree within 10% of mean degree. The V_m proxy is produced by convolving the firing rate of a single neuron with an alpha function with a 4 ms time constant. The top plot shows that with no inhibition (or very little inhibition) activity in this parameter region dies away to zero and is unimodally distributed about a small value. The middle plot shows that moderate amounts of inhibition results in self-sustained activity that is bimodally distributed about one high and one low value. The bottom plot shows that when the fraction of nodes that are inhibitory is much larger than connection density activity is self-sustaining and unimodally distributed about a high value with low variance relative to the mean. **b** shows the scaling relation for the avalanches inferred from $\Phi_i(t)$ at different levels of inhibition, as in panel A. Inhibition detrimentally impacts the validity of the scaling relation predictions, which are required for consistency with

critical branching. The predicted (γ_p) and fitted (γ_p) scaling exponents are indicated as is the goodness of fit (R_p^2) for the predicted exponent. **c** shows how avalanche (fluctuation) statistics vary with the parameter set displayed in panels A and B. The top row shows avalanche (fluctuation) sizes, while the bottom row shows the duration distributions. Exponents τ (size distribution) and β (duration distribution) as well as comparison quotients q are annotated on the plot. From these plots, we can see that temporal smoothing ($\Phi_i(t)$) is necessary to accurately capture $F(t)$. Additionally, we see that mismatch between the $F(t)$ and $P_i(t)$ avalanche distributions vary with network parameters. At high levels of inhibition, the $i(t)$ avalanches are power-law distributed over smaller portions of their support. For $\Phi_i(t)$, neither of the networks with less inhibition show the cut-offs associated with under sampling a critical branching network.

We looked at the magnitude of relative error between estimated exponents for the avalanche size distribution (Figure 2.4c) to determine how well our proxy neural inputs, $\phi_i(t)$, reflected network activity, $F(t)$, in different parameter regions, and with different values for the tuning parameter, λ. Importantly the least error occurred for $\lambda = 1$ with and without the presence of inhibitory nodes. This insensitivity to parameter differences supports the claim (Larremore et al., 2014) that the system becomes critical when $\lambda = 1$ even in the presence of inhibition.

However, the four regions of parameter space perform differently according to our four standardized criteria for consistency with criticality. In the "positive weights" region 90% of 90 trials (nine points in parameter space with ten trials per point) have network activity that meets all four criteria when the tuning parameter is set at criticality ($\lambda = 1$) (Figure 2.4c). Meanwhile 39% meet the criteria in the "ceaseless" region, 19% do in the "quiet" region, and 67% do in the "switching" region which may indicate the location of a second phase-transition and shows that evidence for precise criticality in this model is limited once inhibition is included.

As we vary the tuning parameter, we can clearly distinguish critical from non-critical systems. Over all 47% percent of trials meet all four criteria when $\lambda = 1$, while 3% do when $\lambda = 0.95$, 18% do when $\lambda = 1.015$, 1% do when $\lambda = 0.9$, and 1% do when $\lambda = 1.03$ (Figure 2.4d).

The estimated power-law exponents show that the avalanche size distributions for $F(t)$, $P_i(t)$, and $\Phi_i(t)$ are most alike at criticality. Note that estimated exponents serves as the "scaling index", a measure of the heavy tail even when a power-law is not the statistical model that fits best (Jeżewski, 2004). The fact that matching between network activity and the input integration function was best at criticality is important because it underscores the scale-free nature of critical phenomena and contrasts with the results obtained when testing a different relationship between subsampling methods and network structure (Priesemann et al., 2009; Yu et al., 2014; Levina and Priesemann, 2017).

While the system was both critical ($\lambda = 1$) and in the positive weights region, our V_m proxy $\Phi_i(t)$ met all four criteria for consistency with criticality 74% of the time for 90 trials (Figure 2.4d) while $P_i(t)$ met all four only 1% of the time. The network activity had avalanche size and duration exponent values $\tau_F = 1.43 \pm 0.04$, and $\beta_F = 1.87 \pm 0.09$, (Figure 2.4d) and had a fitted scaling relation exponent, $\gamma_{F_f} = 1.83 \pm 0.02$, and a predicted exponent $\gamma_{F_p} = 1.99 \pm 0.23$. The membrane potential proxy, $\Phi_i(t)$ had slightly lower avalanche size and duration exponent values that fluctuated around the paired network values, $\tau_\Phi = 1.40 \pm 0.06$, and $\beta_\Phi = 1.73 \pm 0.17$, (Figure 2.4d) and exclusively lower scaling relation exponents $\gamma_{\Phi_f} = 1.68 \pm 0.02$. While the unsmoothed $P_i(t)$ varied

considerably more it had size and duration exponents that were almost exclusively higher than the paired network values, $\tau_P = 1.87 \pm 0.50$, and $\beta_P = 2.84 \pm 1.45$, with a fitted scaling relation exponent that was exclusively lower $\gamma_{P_f} = 1.68 \pm 0.02$.

In Figure 2.5, we compared different population dynamics estimation techniques by looking at avalanches inferred from $P_i(t)$ (the inputs to neuron i), and the V_m proxy $\Phi_i(t)$. Both $P_i(t)$ and $\Phi_i(t)$ fluctuate about $F(t)$ but $P_i(t)$ is much noisier (Figure 2.5a), in the ceaseless regime $P_i(t)$ and $\Phi_i(t)$ are systematically offset. Avalanches inferred from $\Phi_i(t)$ had average sizes that scaled with duration (Figure 2.5b). Avalanches from $\Phi_i(t)$ consistently had duration and size distribution exponents that were closer to network avalanches than avalanches from $P_i(t)$. However, $P_i(t)$ performed satisfactorily in the sense that its error was systematically offset and best at criticality (Figure 2.5c).

Including inhibition introduces several important differences. For the ceaseless region with $\lambda = 1$, far fewer trails meet our criteria, however $P_i(t)$ follows $F(t)$ much more closely. The network activity had avalanche size and duration exponent values $\tau_F = 1.48 \pm 0.09$, and $\beta_F = 1.53 \pm 0.09$, and had a fitted scaling relation exponent, $\gamma_{F_f} = 1.23 \pm 0.11$. The membrane potential proxy, $\Phi_i(t)$ had slightly higher avalanche size and duration exponent values that fluctuated around the paired network values, $\tau_\Phi = 1.51 \pm 0.19$, and $\beta_\Phi = 1.57 \pm 0.17$, but nearly identical scaling relation exponents $\gamma_{\Phi_f} = 1.23 \pm 0.11$. While the unsmoothed $P_i(t)$ varied considerably more, it had size and duration exponents that were almost exclusively higher than the paired network values, $\tau_P = 1.88 \pm 0.20$, and $\beta_P =$

2.18 ± 0.34, with a fitted scaling relation exponent that was slightly lower $\gamma_{P_f} = 1.19 \pm 0.07$.

When $\lambda \neq 1$ both $\Phi_i(t)$, and $P_i(t)$ failed to meet all four criteria for criticality at the same high rate as $F(t)$ (to within 1%). This lack of false positives confirms that these signals are useful for characterizing critical branching. In Figure 2.4, panel B, we calculated the absolute magnitude of relative error between the size exponent from avalanche analysis performed on $F(t)$ and $\Phi_i(t)$. As expected, the avalanches were usually not power-laws according to our standards, in this case the exponent is known as the "scaling index" and describes the decay of the distribution's heavy tail (Jeżewski, 2004).

When we set $\lambda = 0.95$ we see a moderate deterioration in the ability of either $\Phi_i(t)$ or $P_i(t)$ to recapitulate network exponent values. The error is no longer systematic; thus, they cannot be used to predict network values. The variability of the exponents increases greatly for $\Phi_i(t)$ while it decreases for $P_i(t)$. The exponent error increases slightly over the $\lambda = 1$ and the base of the distribution is much broader.

Reducing λ further, to $\lambda = 0.90$, the input integration function, $P_i(t) \sim \lambda\omega_i(t-1)$, rapidly dampens impulses (ω_i is the instantaneous firing rate over the presynaptic population for neuron i). Variability continues to increase, and a systematic offset does not return. Exponent error is now much broader. With branching this low, events often are not able to propagate to the randomly selected neuron, an exception is the "ceaseless" regime where activity is still long lived.

When we set $\lambda = 1.015$ we see a dramatic deterioration in the ability of either $\Phi_i(t)$ or $P_i(t)$ to recapitulate network values. Variability in exponent estimation increases for both $\Phi_i(t)$ and $P_i(t)$. Exponent error increases rapidly, underscoring the inability to estimate network activity from neuron inputs.

Increasing λ further to $\lambda = 1.03$ produces an input integration function, $P_i(t) \sim \lambda\omega_i(t-1)$, that rapidly amplifies all impulses and the network saturates. The effect is that variability in the estimated exponents decreases and a systematic offset returns, with both $\Phi_i(t)$ and $P_i(t)$ producing exponents that are exclusively and considerably higher than network values. Exponent error reveals that estimating network properties from the inputs to a neuron is probably not possible for supercriticality in this model.

The results here show that the V_m proxy represents an effective way of subsampling network flow. This is a hallmark of the near-critical region in the PIF model and a manifestation of scale-freeness. Criticality in our model corresponds to the point when the inputs to a neuron represent an average of the activity of the presynaptic population. Importantly we explored why it works, as well as showing that it does work in experimental data. This analysis, presented in forthcoming sections, uncovered that proper temporal and spatial aggregation is important as is the role of inhibition in membrane potential dynamics. This supports both the criticality hypothesis, and tight balance (Barrett et al., 2013; Boerlin et al., 2013; Denève and Machens, 2016). Additionally, it has specific implications for the information content of membrane potential.

2.3.4 The Predicted Scaling Relation Exponent is More Stable than Avalanche Size or Duration Exponents

A key part of the study of criticality in neural systems is the assumption that biological systems must self-organize to a critical point. The precise critical point is a very small target for a self-organizing mechanism in any natural system. So, a key question is whether the self-organizing mechanism of the brain prioritizes efficiently achieving information processing advantages of scale-free covariance at the expense of being slightly sub or super-critical (which is a larger target) (Priesemann, 2014; Tomen et al., 2014; Williams-García et al., 2014; Gautam et al., 2015; Clawson et al., 2017).

Our data offered unexpected insight. It is known that so long as three requirements are met the scaling relation will be marginally obeyed: Avalanche size and durations must be power-law distributed (with exponents τ and β respectively) and average size must scale with duration according to a power-law with exponent γ. Given those three requirements one can derive a prediction for the scaling exponent, $\gamma_p = (\beta - 1)/(\tau - 1)$ without needing to assume criticality (Scarpetta et al., 2018). However, without any other assumptions one expects β and τ to be independent so plotting one against the other should make a point-cloud that is symmetrical, not stretched along a trendline (Figure 2.3).

We analyzed the independence of τ, β, and γ measured from experimental data (where self-organization is hypothesized) and compared it to model data (where self-organization is impossible, but criticality is guaranteed). We found that β and τ are more independent

and the predicted scaling relation is more variable for the model than for experimental data in which β and τ covary, apparently in order to maintain a fixed scaling relation prediction.

The previous multi-site LFP recordings displayed a range of values for the avalanche size τ and duration β distribution exponents across the tested brain preparations. Interestingly, the exponent values were not independent, rather the duration exponent varied linearly with the size exponent (Shew et al., 2015) (Figure 2.3a). The single-neuron V_m fluctuations, reported here, produced a similar linear relationship between size and duration exponent (Figure 2.3b). Algebraic manipulation of the predicted scaling exponent $\gamma_p = (\beta - 1)/(\tau - 1)$ provides a clue. If the scaling relation $(\beta - 1) = \gamma(\tau - 1)$ is obeyed and if γ_p is a *fixed universal property,* then the linear relationship $\beta_j \propto \langle\gamma_p\rangle\tau_j$ holds across different cells and animals.

To demonstrate this important result, variability in the predicted scaling-relation is much less than expected, we propagate errors and assume independent β and τ. We would expect the standard deviation of γ_p to be $\sigma^*_{\gamma_p} \sim \frac{|\beta-1|}{|\tau-1|}\sqrt{\left(\frac{\sigma_\beta}{\beta-1}\right)^2 + \left(\frac{\sigma_\tau}{\tau-1}\right)^2} \sim 0.72$ which is roughly twice the real value in V_m data, $\sigma_{\gamma_p} \sim 0.35$.

Pearson correlation, ρ, confirms strong dependence between τ and β, $\rho_{\tau\beta} = 0.61$, p-value $p = 2.57 \times 10^{-6}$ for the V_m data while for the MEA data $\rho_{\tau\beta} = 0.96$, p-value $p = 1.01 \times 10^{-7}$. From this we confirm what Figure 2.3 shows: the variability in τ and β are not independent and this implies the existence of an organizing principle connecting τ to β.

Whatever the principle may turn out to be, one of its effects is the maintenance of low variability in γ_p at the expense of greater variability in τ and β.

A principal reason to suspect self-organization is that this trend is not seen in the model results. Importantly, τ and β are independent of the scaling-relation exponent function, though still weakly correlated. In this model there is no adaptive organizing principle driving this network to criticality, instead the structure is fixed and set to be at the critical point. This shows how systems behave in the absence of self-organization. No parameter is being maintained at low variability at the expense of other parameters.

Limiting ourselves to simulated network activity for the $\lambda = 1$ case without inhibition (Figure 2.4c), propagation of errors leads us to expect the standard deviation of the scaling-relation prediction to be $\sigma^*_{\gamma_{F_p}} \sim 0.27$ which is very close to real value $\sigma_{\gamma_{F_p}} \sim 0.23$. The correlation is statistically significant at the 5% level, but much smaller $\rho_{\tau\beta} = 0.23$, p-value $p = 0.027$.

This was noted in the original paper (Shew et al., 2015) where they were able to reproduce the linear trend between avalanche size and duration exponents by simulating a network with synaptic depression to adaptively restore critical behavior after an increase in network drive. They show that the trendline is produced by corrupting their simulated data via randomly deleting seventy to ninety percent of spiking events and then changing the way they group events in time (adaptive time binning). Our membrane potential

fluctuations have no counterpart to the adaptive time binning other than the intrinsic membrane time constant which is not manipulated experimentally.

In conclusion, the linear trend between avalanche size and duration exponents is not a universal property of critical systems because it was not found in the model. This suggests that the linear trend is enforced by an organizing principle at work in the brain but absent in the model. This principle prioritizes maintaining stability in either the scaling of avalanche size with duration, or the power-law scaling of autocorrelation which is closely related to the scaling relation and scale-free covariance via the power-law governing auto-correlation (Bak et al., 1987; Sethna et al., 2001).

2.3.5 Non-Linearity and Temporal Characteristics such as High-Order Correlation, Proper Combination of Synaptic Events, and Signal Time-Scale are Required to Reproduce Network Measures from Single-Electrode Recordings

In order to demonstrate that subthreshold membrane potential fluctuations can be used as an informative gauge of cortical population activity it is necessary to compare against alternative signals which have either been used by experimentalists as a measure of population activity or that share some key features of membrane potential but are missing others. By making these comparisons we can illuminate which features of the membrane potential signal are responsible for its ability to preserve properties of cortical network activity. Additionally, it is necessary to check whether the statistical properties of avalanches can be explained by random processes unrelated to criticality. To address these points of the investigation, we analyzed five surrogate signals: single-site LFP recorded

concurrently with the V_m recordings, two phase-shuffled versions of V_m recordings, computationally inferred excitatory current, and the same inferred excitatory current further transformed to match V_m autocorrelation (which tests the role of IPSPs by making a V_m-like signal that lacks them).

2.3.6 Negative Fluctuations of LFP Disagree with V_m and MEA Results and are Inconsistent with Avalanches in Critical Systems

The first alternative hypothesis to test is whether the LFP could yield the same results. We used low-pass filtered and inverted single site local field potential (LFP) which is commonly believed to measure local population activity. However, in our analysis it did not recapitulate the results from either MEA or V_m avalanche analysis. We obtained viable single-site LFP recordings (see Methods: Extracellular Recordings), simultaneous and adjacent with whole-cell recordings, for 38 of the 51 neurons reported above. We performed avalanche analyses on the LFP recordings using a procedure like the one described for the V_m recordings (see Methods: Intracellular Recordings) (Figure 2.6). LFP recordings were grouped the same way V_m recordings were in order to match them for comparison. However, the numbers of recordings are not the same because there were two or three cells being patched alongside (within 300 μm) one extracellular electrode and there was not always a simultaneous LFP recording. LFP also produced more avalanches per 2-5-minute recording $N_{AV} = 1128 \pm 348$. The are 23 20-minute periods spanning multiple LFP recordings. These recordings were gathered into groups and matched against 49 V_m recording groups (38 from the first 20-minute period, 11 from the second). Additionally, there were 16 20-minute periods spanning only one LFP recording but with

more than 500 avalanches. The concurrent V_m recordings did not have enough avalanches. This gives us 39 LFP avalanche data sets.

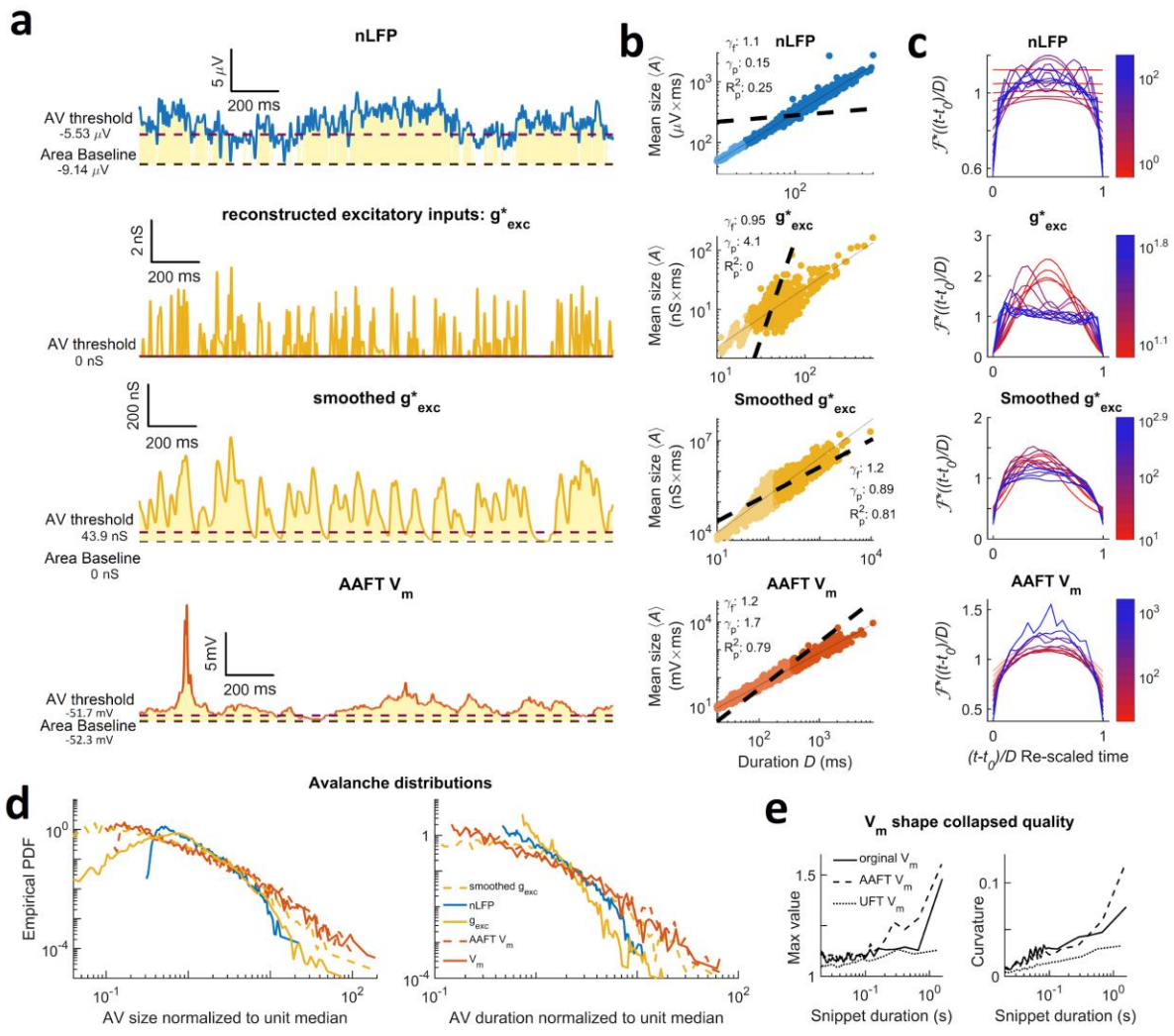

Figure 2.6 | Comparison to surrogate signals reveals the importance of non-linearity and temporal characteristics such as high-order correlation, proper combination of synaptic events, and signal time-scale. a shows alternative signals and surrogate data time synchronized to Figure 2.2b and showing thresholds and integration baselines (dashed lines) with avalanche areas marked in yellow. The top row shows the inverted LFP signal. The LFP is low-pass filtered (0-100 Hz), inverted, detrended and analyzed for avalanches identically to membrane potentials. The second and third rows show the inferred excitatory inputs to a neuron. An algorithm reconstructs the timing and shape of ePSPs from V_m. The resultant signal, g^*_{exc}, is much faster, making it analogous to the $P_i(t)$ signal from the PIF model. This signal is smoothed (third row, see Methods: Model Simulations for details) to produce a signal that is like V_m (Figure 2.2b) would be if it lacked IPSPs. The last row provides an example of amplitude matched phase shuffled surrogate data (amplitude adjusted Fourier transform algorithm). **b** shows the scaling relation in the same order and dataset as panel A. The dashed line is the predicted scaling relation exponent inferred from

power-law fits to the size and duration distributions of positive fluctuations. In cases where a power-law is not the best model the exponent nonetheless gives the average slope of a linear regression on a log-log plot, a "scaling index" (Jeżewski, 2004). The predicted (γ_p) and fitted (γ_f) scaling exponents are indicated as is the goodness of fit (R_p^2) for the predicted exponent. Mean size scales with duration for all signals but often it is trivial ($\gamma_f \sim 1$) or poorly explained by a power-law ($R_f^2 < 0.95$), and it is rarely a good match with the prediction from the scaling relation. **c** shows shape collapse from the total dataset in the same order and dataset as panel A. The color indicates the duration according to the scale bar. If self-similarity is present each avalanche profile will collapse onto the same curve: $F\left(\frac{t-t_0}{D}\right)$. The LFP illustrates a trivial scaling relation that is not produced by true self-similarity: limited curvature and the exponents are very close to one. The second row shows the reconstructed excitatory inputs, g_{exc}^* , and lacks shape collapse as expected from the lack of a scaling relation power-law in panel B. The third row shows that sensible curvature re-emerges with smoothing but does not produce a universal scaling function. In the last row the phase shuffled V_m shows a shape collapse which is worse than for the original V_m (Figure 2.2e). **d** shows size and duration distributions from each signal compared with the V_m (in solid red). The phase shuffled V_m (dashed red) still obeys power-laws but the exponent values disagree, and it less frequently meets our standardized criteria. Unsmoothed g_{exc}^* (solid gold) is more like inverted LFP than anything else. When g_{exc}^* is smoothed (dashed gold) it becomes closer to the original V_m but retains pronounced curvature in the duration distribution. We see V_m, AAFT, and smoothed g_{exc}^* produce distributions which extend over similar orders of magnitude ($\sim$2). **e** shows maximum value and curvature of the average profiles after "collapse" as functions of duration. Shape collapse quality is a subjective measure, but these give a more quantitative perspective. Good shape collapse should have a fixed maximum value and a high but fixed mean curvature. For comparison, the UFT (Unwindowed or Unadjusted Fourier Transform) phase shuffled data is also shown to provide a comparison to low curvature but a fixed maximum value. By visual inspection of AAFT and V_m it is apparent that the asymmetry is gone and that deviation from the collapsed shape begins at shorter durations. The max value diverges from a linear trend sooner for AAFT ($\sim$0.15 seconds, 0.5) than for V_m ($\sim$0.7 seconds). Curvature also diverges sooner for the AAFT (0.5 seconds vs 0.7 seconds). Curvature does not become appreciable until about 50-70 ms. Between the onset of curvature and divergence of max value there are $\left(\frac{0.15}{0.05}\right) \sim 0.48$ orders of magnitude for AAFT and $\left(\frac{0.7}{0.05}\right) \sim 1.15$ orders of magnitude for the original V_m.

The LFP recording groups performed poorly according to our four criteria for consistency with criticality. Of the 39 LFP recording groups, only 41% percent had acceptable scaling relation predictions and only 36% met all four standard criteria for criticality (Figure 2.7a). The additional criterion of shape collapse was not observed (Figure 2.6c), there was no linear trend among the exponents governed by the scaling relation and the exponents did not match MEA data (Figure 2.3a). However, 85% produced power-law fits for size and duration, 92% had scaling relations well fit by power-laws and all were non-trivial. We expect from (Touboul and Destexhe, 2017) that some fraction of non-critical data will pass the four standard criteria by chance, so long as the data have a $1/f$ power spectrum.

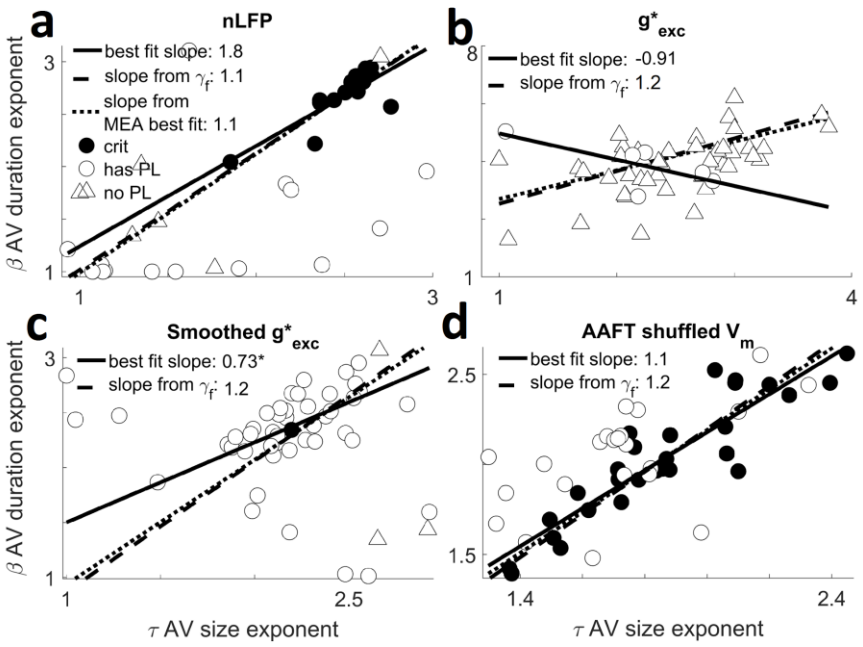

Figure 2.7 | Plausible alternative signals fail to demonstrate consistency with criticality. A plot of the exponents governing power-law scaling of avalanche duration vs the exponents governing avalanche size. Circles indicate data which was best fit to a power-law in both its size and duration. Triangle indicates otherwise (the MLE estimation of a would-be power-law fit, the "scaling index", is plotted in that case (Jeżewski, 2004)). Filled circles indicate data that meet all four standardized criteria for judging data to be consistent with criticality. We show the performance summary for the first group of data from each cell (the first 20-minute period which contained multiple recordings). The best fit slope is from linear regression to the plotted or indicated data, this is compared to the slope predicted by the mean γ_f (the exponent describing how avalanche size scales with duration). **a** shows that positive fluctuations of inverted LFP were less likely to be power-law distributed and the power-law exponents tended to be unstable and not resemble MEA results. All 39 LFP datasets are represented. **b** shows results from the reconstruction of excitatory input conductance g_{exc}. Only 12% were power-law distributed. The results do not resemble the MEA results. The slope from the trendline matches the scaling relation exponent but the regression is bad, $R^2 = 0.51$. **c** shows how adding back some temporal smoothing to g^*_{exc} can improve results, 94% have power-laws but the exponents are more variable and generally larger. Most (96%) fail to have scaling relations which are well described by power-laws. The exponents β and τ are less independent but are not well described by the regression trendlines ($R^2 = 0.35$). The fit is applied only to the upper right cluster, excluding the outliers in the region $\beta < 1.6$ and $\tau < 1.6$. **d** shows the summary of results from the AAFT phase shuffled V_m. As expected for a shuffling that preserves autocorrelation, power-laws are also preserved. However, the exponents are shifted down (especially the size distribution exponent). Far more fail to meet our criteria for consistency with criticality, as statistically significant difference (see Results: Stochastic Surrogates are Distinguishable from Vm or MEA Results, Reveals Importance of Non-Linear Filtering). Significantly fewer data sets have scaling relations well described by a power-law (75% as opposed to 90%), this is consistent with a slightly worse shape collapse (Figure 2.6c).

To emphasize that these results are chance we can limit ourselves to just those with the best chance of meeting the scaling relation criteria by picking those that have power-laws in the size and duration distributions. This is enough to expect the scaling relation to be

obeyed if mean size scales geometrically with duration (Scarpetta et al., 2018). It is still the case that only 42% of recording groups meet the three remaining standard criteria for consistency with criticality. Therefore, having power-laws is statistically independent of meeting the other criterion for consistency with criticality.

Not only does the single-site LFP data differ from MEA and V_m data because it fails to demonstrate consistency with criticality, it is also the case that the scale-free properties which do exist are not representative of the MEA data or the simultaneous V_m recordings. The failure was not because LFP recordings co-occurred with decreased consistency with criticality more generally. Eighty-one percent of the matched V_m recordings met all the criteria, while 58% of the LFP recordings did, a statistically significant dissimilarity, odds ratio ($r_{OR} = 7.65$ with $p = 1.1 \times 10^{-5}$).

The estimated exponents from all 39 LFP recording groups were highly variable. The duration distribution and scaling relation were most dissimilar to V_m and MEA data. Of the 33 LFP groups which were power-law distributed, the avalanche size exponent had a median value $\tau = 1.90 \pm 0.63$ while the duration exponent was $\beta = 1.41 \pm 0.9$ (very low) (Figure 2.7a) and the fitted exponent was $\gamma_f = 1.11 \pm 0.02$. The predicted scaling-relation exponents were inaccurate with $\gamma_p = 0.89 \pm 0.76$ for the subset of recording groups which had power-laws.

The extreme variability makes it hard to determine whether the size and duration exponents match other data, but the fitted scaling relation exponent was much less variable

and more clearly separated from MEA or V_m results. The matched difference of median test (Wilcoxon signed-rank) between 49 recording groups found that the best fit τ ($\tau = 1.90 \pm 0.63$) was not significantly distinguishable from the Vm data ($r_{SDF} = 0.15, p = 0.33$), but β ($\beta = 1.41 \pm 0.9$) was dissimilar with a comparable effect size ($r_{SDF} = 0.17, p = 0.028$), and γ_f ($\gamma_f = 1.11 \pm 0.02$) was also dissimilar ($r_{SDF} = 0.25, p = 7.1 \times 10^{-15}$).

When comparing to the 13 samples of MEA data γ_f was significantly different from the MEA data ($r_{SDF} = 0.88$, and $p = 9.21 \times 10^{-08}$). This contrasts with our comparison between V_m and MEA data. In that case the scaling relation was not distinguishable even with 51 points of comparison and very low variability making a difference easier to detect. However, because of their extreme variability the size and duration exponents fail a 5% significance threshold for distinguishing from the MEA data by a Wilcoxon rank-sum result ($r_{SDF} = 0.06, p = 0.766$ for τ and $r_{SDF} = 0.29, p = 0.123$ for β). This failure of inverted LFP to show the same statistical properties as multi-unit activity may add a caveat to the assumptions behind the use of inverted LFP as a proxy for population activity (Kelly et al., 2010; Einevoll et al., 2013; Okun et al., 2015). Specifically, the amplitude of *single-electrode* negative LFP excursions is ambiguously related to the number of spiking neurons, whereas the use of electrode arrays as in (Beggs and Plenz, 2003) and in (Shew et al., 2015) is more appropriate.

To summarize, the single-site LFP fluctuation results from the superposition of local spiking and extracellular synaptic current of juxtaposed network elements (Kajikawa and Schroeder, 2011; Einevoll et al., 2013; Pettersen et al., 2014; Ness et al., 2016). These

fluctuations were found to be less informative about the network dynamics than single-neuron V_m fluctuations. V_m fluctuations result from the superposition of EPSPs and IPSPs indicating neuronal responses propagating in a manner consistent with the true neural network architecture. In other words, synaptic and spiking events driving fluctuations at single extracellular electrodes may be too badly out of sequence and distorted to faithfully represent neuronal avalanches, whereas the sequence of synaptic and spiking events driving somatic membrane potential fluctuations is functionally relevant by definition.

2.3.7 Stochastic Surrogates are Distinguishable from V_m or MEA Results, Reveal Importance of Non-Linear Filtering

After eliminating inverted LFP as an alternative single-electrode signal, it was important to establish whether our results could have been created from a linear combination of independent random processes (Touboul and Destexhe, 2017; Priesemann and Shriki, 2018) , similar to those used when contesting evidence for critical brain dynamics (Bédard et al., 2006; Touboul and Destexhe, 2010, 2017). We also wanted to learn what effects non-linearity (non-Gaussianity) has in signals like the V_m.

To address these questions, we used both the AAFT and UFT phase shuffling algorithms (see Methods: Experimental Design and Statistical Analysis). AAFT (Figure 2.6) preserves both the exact power-spectrum (autocorrelation) of the signal and non-linear skew of signal values but randomizes the phase (higher-order temporal correlations). UFT is the same but forces the distribution of signal values to be Gaussian. Using both allows us to

attribute some characteristics to non-linear rescaling and others to precise temporal correlation structure.

Phase shuffling tends to preserve power-laws since it explicitly preserves the 1/f trend of the power-spectrum. However, the matched signed-rank test reveals that the values of the exponents change in both methods. Under UFT transformation the scaling relation and shape collapse became more trivial and like the LFP. This suggests that both the non-linear rescaling of input currents by membrane properties and the way that input populations interact throughout the intricate dendritic arborization are important.

For the 51 recording groups from the first 20-minutes the AAFT reshuffled data yield a median size exponent of $\tau = 1.74 \pm 0.29$ while the duration exponent was $\beta = 2.0 \pm 0.34$ (Figure 2.7d). The fitted scaling relation exponent was $\gamma_f = 1.19 \pm 0.06$ and the predicted scaling relation exponent was $\gamma_p = 1.21 \pm 0.49$.

Pairing the surrogates to the original V_m data and performing the Wilcoxon signed-rank test for difference of medians gives ($r_{SDF} = 0.053,\ p = 2 \times 10^{-4}$), ($r_{SDF} = 0.091,\ p = 0.08$), and ($r_{SDF} = 0.207, p = 3 \times 10^{-5}$) for τ, β, and γ_f respectively. Thus τ and γ_f are both significantly different, this is supported by the fact that only 55% of the groups meet all four standard criteria for criticality, while 76% of meet them for the original V_m time series. This difference between success rates is significant by Fisher's exact test ($r_{OR} = 2.67, p = 0.0363$).

The failure mode for AAFT shuffled data was almost entirely in reduced goodness of fit (R^2) for a power-law fit to its scaling relation, 17% fewer recording groups met the criterion $R^2 > 0.95$, than for V_m ($r_{OR} = 4.18, p = 0.0093$). When the shape collapse is examined, we see another clear, if qualitative, difference in the symmetry of any presumed scaling function (Figure 2.6c). When taken together can we see that the AAFT shuffled dataset is not consistent with critical point behavior. Thus, we show that the exponent values and evidence for criticality, especially scaling and shape collapse which we inferred from V_m are not likely to come from random processes and are dependent on non-linear temporal correlation structure.

The key feature of the UFT result is that the fitted scaling relation exponent is much lower, $\gamma_f = 1.05 \pm 0.049$, which is significantly less than for AAFT ($r_{SDF} = 0.25, p = 1 \times 10^{-13}$) and less than the LFP ($r_{SDF} = 0.228, p = 3 \times 10^{-6}$). It is very close to trivial scaling but is still distinguishable from $\gamma_f = 1$ at a population level via the sign test ($r_{SDF} = 0.843, p = 2 \times 10^{-10}$). Because the fitted scaling relation exponent and shape collapse were similar in both the UFT and LFP data, it suggests that lack of non-linear rescaling (non-linear filtering) may be a key feature of LFP that explains its failure to accurately reflect critical point behavior.

The UFT was universally poorer performing, 39% do pass the criticality test but given that the scaling relation exponent is so low this is simply random chance, and significantly worse than the V_m results ($r_{OR} = 5.04, p = 3 \times 10^{-4}$). The UFT phase shuffling results obtain a median size exponent of $\tau = 1.69 \pm 0.45$ while the duration exponent was $\beta =$

1.81 ± 0.49. The predicted scaling relation exponent was $\gamma_p = 1.01 \pm 0.72$. All are significantly different from the V_m results ($r_{SDF} = 0.183,\ p = 0.005$), ($r_{SDF} = 0.199,\ p = 2 \times 10^{-4}$), and ($r_{SDF} = 0.249, p = 2 \times 10^{-13}$) for τ, β, and γ_f respectively. These results are redundant with the AAFT confirming that our results do not have a trivial explanation.

When the scaling relation was examined, we saw another clear, if qualitative, difference in the symmetry of any presumed scaling function (Figure 2.6c). When taken together, our four standardized criteria followed by shape-collapse analysis let us distinguish phase-shuffled V_m fluctuations from the original V_m fluctuations, even limiting ourselves to data that meets the four criteria. Thus, the phase-shuffled data showed that the evidence for criticality in the original V_m fluctuations are dependent on non-linear temporal correlations.

2.3.8 Excitatory and Inhibitory Synaptic Activity are both Required for V_m Fluctuations to Match MEA Avalanches

Having learned that single-site LFP recordings cannot be used to accurately infer the statistics of population activity, and knowing that low-pass filtered and inverted LFP is believed to reflect *excitatory* synaptic activity (Kajikawa and Schroeder, 2011; Buzsáki et al., 2012; Einevoll et al., 2013; Ness et al., 2016) it begs the question: to what extent do excitatory synaptic events contain evidence for network criticality?

Somatic V_m fluctuations are the complex result of spatially and temporally distributed excitatory and inhibitory synaptic inputs further mangled by active and passive membrane

properties in dendrites and soma. There is reason to believe that these features conspire to enforce the condition that V_m faithfully represents inputs to the presynaptic network (Barrett et al., 2013; Boerlin et al., 2013; Denève and Machens, 2016) similar to how input signals relate to presynaptic populations in our model. To address the stated question, we estimated the excitatory synaptic conductance changes g^*_{exc} from the V_m recordings, using a previously developed inverse modeling algorithm (Yaşar et al., 2016), and applied the avalanche analysis on the inferred g^*_{exc} time series, (Figure 2.6).

The inferred excitatory conductance is plausibly related to the presynaptic population, however it failed to be a reliable measure of network dynamics (Figure 2.7b). We can't know whether the failure is because excitatory current does not contain enough information or because the signal's time constant is too short. Power laws in the avalanche size and duration distributions were observed in only 12% of the 51 groups from the first 20 minutes of recording. Comparing to V_m this was very different ($r_{OR} = 375, p = 6 \times 10^{-14}$). Shape collapse was absent from the inferred excitatory conductance (Figure 2.6c) and none passed all four criteria for criticality. From this we conclude that inferred excitatory conductances are not a good network measure.

One of many potential reasons for this failure could be the much shorter time constant of the inferred g^*_{exc} signal compared to the V_m signal. We saw exactly that situation when examining model results: $P_i(t)$ failed to reproduce network values as well as its smoothed version $\phi_i(t)$. Therefore, we smoothed the g^*_{exc} signal with an alpha-function, chosen because it should impose a similar non-Gaussian distribution as the V_m signal. The time

constant of the alpha function was tuned to minimize the error between the autocorrelation of the smoothed g^*_{exc} signal and the original V_m signal. By doing so we create a signal with a $1/f$ power-spectrum that should exhibit power-laws and reproduce many V_m statistical features, (Figure 2.6).

Reinstating the autocorrelation does not summon the return of scale-freeness. The smoothed signal did demonstrate power-laws (94%) and one serendipitously met the standardized criteria for consistency with critical point behavior (Figure 2.6d). However, this is chance. The average coefficient of determination for a fitted scaling relation on a log-log plot was $R^2 = 0.84 \pm 0.14$ so overall average avalanche sizes did not scale with duration as a power-law. Nonetheless this is a substantial improvement on the unsmoothed version $R^2 = 0.68 \pm 0.17$. This is a statistically significant difference ($r_{SDF} = 0.054, p = 3 \times 10^{-4}$).

The smoothed inferred g^*_{exc} signal (Figure 2.6a) is visually more like the original V_m (Figure 2.2b) than the AAFT shuffled V_m surrogate (Figure 2.6a), however, it was a worse match. This shows that signals dependent only on excitation, even ones with the same non-Gaussian distribution and power-spectrum trend do not reflect the statistics of population activity. Interactions between EPSPs and IPSPs may be needed.

In conclusion, the single-site local field potential (LFP), the phase-shuffled recorded V_m, and the inferred excitatory conductance g^*_{exc}, including its smoothed version, all failed to reveal the critical network dynamics. However, there are either similarities between the

signals or some remaining scale-free signatures which reveal the importance of signal aspects. In order to faithfully represent population activity statistics a candidate signal must: have the right non-Gaussian distribution, the right $1/f$ power-spectrum characteristics and is sensitively dependent on higher-order temporal correlations such as may result from the complex interplay of excitation and inhibition within the dendritic arborization of a pyramidal neuron in the visual cortex.

2.4 Discussion

Leveraging membrane potential (V_m) and local field potential (LFP) recordings with modeling and microelectrode array (MEA) data yielded two principal findings: subthreshold V_m are a useful indicator of network activity and this correspondence is inherent to critical coarse-graining. Scrutiny revealed that avalanche size and duration distribution parameters covary to maintain similar geometrical scaling across different experiments, a noteworthy observation. The following discussion emphasizes possible significance and research intersections, such as explaining disagreement with theory via subsampling effects or quasicriticality, or relating neural computation to a mathematical apparatus within critical systems theory.

While “appropriating the brain’s own subsampling method” is a novel description of whole-cell recordings, it was inspired by examples. Whole-cell recordings contain information about the network (Gasparini and Magee, 2006; Mokeichev et al., 2007; Poulet and Petersen, 2008; El Boustani et al., 2009; Okun et al., 2015; Malina et al., 2016; Hulse et al.,

2017; Lee and Brecht, 2018) and stimulus (Anderson et al., 2000; Sachidhanandam et al., 2013). Usually the focus is using neural inputs to predict outputs, not measuring population dynamics (Destexhe and Paré, 1999; Carandini and Ferster, 2000; Isaacson and Scanziani, 2011; Okun et al., 2015). Additionally, long-time or large-population statistics, like our avalanche analysis, are useful for understanding neural code (Sachdev et al., 2004; Churchland et al., 2010; Crochet et al., 2011; Graupner and Reyes, 2013; McGinley et al., 2015; Gao et al., 2016) and are robust to noise. Our finding that single V_m recordings reflect scale-free network activity is significant as recording stability in behaving animals improves (Poulet and Petersen, 2008; Kodandaramaiah et al., 2012; Lee and Brecht, 2018). We open the door to using V_m fluctuations as windows into network dynamics.

Rigorous analysis supports our experimental conclusion: subthreshold V_m fluctuations mimic neuronal avalanches and evince critical phenomena but negative LFP deflections don't, despite being purported network indicators (Bédard et al., 2006; Liu and Newsome, 2006; Kelly et al., 2010; Einevoll et al., 2013; Okun et al., 2015). We invoke network not single-neuron criticality (Gal and Marom, 2013; Taillefumier and Magnasco, 2013) because the trend between size and duration exponents agrees with MEA data. Our findings originate from spontaneous activity of ex-vivo turtle visual cortex which shares many connectivity and functional features with mammalian cortex (Ulinski, 1990; Larkum et al., 2008). Lastly, the results are not serendipitous noise because the V_m dataset significantly differed from a dataset of phase-shuffled and rescaled surrogates (Theiler et al., 1992).

Readers keen on critical phenomena may notice our exponents differ from the exact theoretical predictions ($\tau = 1.5, \beta = 2$ (Haldeman and Beggs, 2005)). Others observing this mismatch have suggested the brain operates slightly off-critical (Hahn et al., 2010; Priesemann, 2014; Tomen et al., 2014).

An extension of this suggestion, quasicriticality (Williams-García et al., 2014), also explains the highly stable scaling relation: biological systems blocked from precise critically may optimize properties which are maximized only for critical systems, becoming "quasicritical". Correlation time and length are maximized only at criticality and closely related to avalanche geometrical scaling (Tang and Bak, 1988; Sethna et al., 2001). If brains optimize correlation length, a highly stable scaling relation may result. Furthermore, including inhibition (Larremore et al., 2014) makes our otherwise critical model less consistent with criticality except that population statistics can still be inferred from input fluctuations. The stable-scaling was not in the model, which lacks any plasticity mechanisms. Stable-scaling may be a rare observation of self-organization principles such as quasicriticality. A contributing explanation is subsampling effects (Priesemann et al., 2009; Levina and Priesemann, 2017) but it doesn't explain the stable scaling relation unless quasicriticality is also invoked.

2.4.1 Neuronal Avalanche and Neural Input Fluctuation Similarity is Captured by a Critical Recurrent Coarse-Graining Network

Our main modeling finding, inputs to a neuron reflect network activity best for critical branching networks, is supported by a parameter sweep and detailed analysis. Our

network had no structure, but structure exists at all scales of brain networks (Song et al., 2005; Perin et al., 2011; Shimono and Beggs, 2015) and can have profound impacts on network dynamics (Litwin-Kumar and Doiron, 2012; Mastrogiuseppe and Ostojic, 2018). We derived a relationship showing that the findings may be transferrable to networks where neural inputs fluctuate about proportionality to some subsample's activity. We tune proportionality to be one, but that can also emerge from plasticity (Shew et al., 2015; Del Papa et al., 2017). Tight-balance suggests a biological mechanism causing subthreshold V_m to track excitation into a presynaptic population because IPSPs can have their timing and strength "balanced" to truncate EPSPs which would otherwise last longer than spurts of presynaptic excitation (Barrett et al., 2013; Boerlin et al., 2013; Gatys et al., 2015; Denève and Machens, 2016). We use V_m proxy, $\phi_i(t)$, an alpha function convolved with a point process, $P_i(t)$. This $\phi_i(t)$, is more like V_m than $P_i(t)$ and reproduces our experimental findings. Lastly, we investigate quasicriticality by including inhibition but tuning the maximum eigenvalue to what would be the critical point without inhibition.

Our model provides insights on network subsampling and renormalization group. Usually subsampling means selecting neurons at random or modeling an MEA with an arbitrary grid (Priesemann et al., 2009). Our "subsample" is the presynaptic population represented by summing weighted inputs from active neurons. This is the first analysis intersecting network convergence (i.e. postsynaptic soma).

Subsampling distorts avalanche size and duration, likely creating differences between experimental results and theoretical predictions (Priesemann et al., 2009; Ribeiro et al.,

2014; Levina and Priesemann, 2017; Wilting and Priesemann, 2018). Subsampling may explain disagreement between avalanche analysis on simulated network activity, $F(t)$, V_m proxy $\phi_i(t)$, and single-neuron firing rate $P_i(t)$. However, V_m and MEA results are off theory but match each other. Either their subsampling errors are alike enough to produce similar distortions, or subsampling co-occurs with quasicriticality (Priesemann, 2014; Williams-García et al., 2014).

Intriguingly, the Restricted Boltzmann Machine (RBM) (Aggarwal, 2018), (a related model) was exactly mapped to a "renormalization group" (RG) operator (Mehta and Schwab, 2014; Koch-Janusz and Ringel, 2018). RG is a mathematical apparatus relating bulk properties to minute interactions (Maris and Kadanoff, 1978; Nishimori and Ortiz, 2011; Sfondrini, 2012). It characterizes critical points of phase-transitions (Stanley, 1999; Sethna et al., 2001) and helps derive neuronal avalanche analysis predictions (Sethna et al., 2001; Le Doussal and Wiese, 2009; Papanikolaou et al., 2011; Cowan et al., 2013). RG operators coarse-grain and then rescale, like resizing a digital image. Crucially, iterating an appropriate operator on a critical system produces statistically identical "copies", but on non-critical systems the iterations diverge. Our model averages (coarse-grains) presynaptic pools to get an instantaneous firing probability for each neuron. Then a logical operation (rescaling) sets the spiking states for the next iteration, demonstrating an RG-like operation that reproduces our experimental findings. Denève and Machens (2016) proposed a similar relationship between real V_m and presynaptic pools. The finding that a similar neural operation emerges in RBMs underscores the relevance of RG and the extension of our findings to structured or non-branching networks. The importance is that

a recurrent coarse-graining network may be like a scale-free ouroboros, displaying widespread scale-freeness if any component is critical or briefly driven by critical or scale-free inputs (Mehta and Schwab, 2014; Schwab et al., 2014; Aoki and Kobayashi, 2017; Koch-Janusz and Ringel, 2018).

Significantly, associating neuronal processing with critical branching may induce an organizing principle, the "Information Bottleneck Principle". This balances dimensionality reduction (compression) against information loss (Tishby and Zaslavsky, 2015) and is reminiscent of efficient coding (Friston, 2010; Denève and Machens, 2016), and origins of tuning curves (Wilson et al., 2016; Heeger, 2017). Koch-Janusz and Ringel (2018) trained their network by maximizing mutual information between many inputs and few outputs. This produced nodes with receptive fields matching popular RG operators. They derived correct power-laws by iterating the network. Applications of RG to neural computation are known: image processing (Gidas, 1989; Mehta and Schwab, 2014; Saremi and Sejnowski, 2016), brain and behavior (Freeman and Cao, 2008), emergent consciousness (Werner, 2012; Fingelkurts et al., 2013; Laughlin, 2014), and hierarchical modular networks (Lee et al., 1986; Willcox, 1991) important for criticality (Moretti and Munoz, 2013). Furthermore, our model's RG-like features are crucial to reproducing experimental results. It follows that elegant RG operators like in the RBM might also capture biological neuronal processing, fulfilling the demand for beautiful neuroscience models (Roberts, 2018) while offering insights into organizing principles and scale-freeness.

2.4.2 Conclusion

We established that subthreshold fluctuations of V_m in single neurons agree with neuronal avalanche statistics and with critical branching but fluctuations in other single-electrode signals do not. Computational modeling showed that accurate inference requires critical branching like connectivity. Fluctuation size scales with duration more self-consistently in experimental than model results, hinting at self-organization. These findings are consistent with a nascent reduction of neural computation to coarse-graining operations which may explain the prevalence of critical-like behavior during spontaneous neural activity. Fully articulating the implications requires more investigation, but we have substantially extended the evidence for critical phenomena in neural systems while rigorously demonstrating that subthreshold V_m fluctuations of single neurons contain useful information about dynamical network properties.

References

Aggarwal CC (2018) Neural Networks and Deep Learning: A Textbook. Springer International Publishing. ISBN: 978-3-319-94462-3.

Anderson J, Lampl I, Reichova I, Carandini M, Ferster D (2000) Stimulus dependence of two-state fluctuations of membrane potential in cat visual cortex. Nat Neurosci 3:617–621 doi: 10.1038/75797.

Aoki K-I, Kobayashi T (2017) Restricted Boltzmann Machines for the Long Range Ising Models. Mod Phys Lett B 30:1650401 doi: 10.1142/S0217984916504017.

Arviv O, Goldstein A, Shriki O (2015) Near-Critical Dynamics in Stimulus-Evoked Activity of the Human Brain and Its Relation to Spontaneous Resting-State Activity. J Neurosci 35:13927–13942 doi: 10.1523/JNEUROSCI.0477-15.2015.

Bak P, Tang C, Wiesenfeld K (1987) Self-organized criticality: An explanation of the 1/f noise. Phys Rev Lett 59:381–384 doi: 10.1103/PhysRevLett.59.381.

Barrett D, Denève S, Machens C (2013) Firing rate predictions in optimal balanced networks. In: Advances in Neural Information Processing (NIPS) (Pereira F, Burges CJC, Bottou L, Weinberger KQ, eds), pp 1–9. Curran Associates, Inc. ISBN: 9781632660244.

Bédard C, Kröger H, Destexhe A (2006) Does the 1/f frequency scaling of brain signals reflect self-organized critical states? Phys Rev Lett 97:118102 doi: 10.1103/PhysRevLett.97.118102.

Beggs JM (2008) The criticality hypothesis: How local cortical networks might optimize information processing. Philos Trans R Soc A Math Phys Eng Sci 366:329–343 doi: 10.1098/rsta.2007.2092.

Beggs JM, Plenz D (2003) Neuronal Avalanches in Neocortical Circuits. J Neurosci 23:11167–11177 doi: 10.1523/JNEUROSCI.23-35-11167.2003.

Beggs JM, Timme N (2012) Being critical of criticality in the brain. Front Physiol 3 JUN:163 doi: 10.3389/fphys.2012.00163.

Bender R, Lange S (2001) Adjusting for multiple testing - When and how? J Clin Epidemiol 54:343–349 doi: 10.1016/S0895-4356(00)00314-0.

Boerlin M, Machens CK, Denève S (2013) Predictive Coding of Dynamical Variables in Balanced Spiking Networks. PLoS Comput Biol 9:e1003258 doi: 10.1371/journal.pcbi.1003258.

Bozdogan H (1987) Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. Psychometrika 52:345–370 doi: 10.1007/BF02294361.

Brunel N, Hakim V, Richardson MJE (2014) Single neuron dynamics and computation. Curr Opin Neurobiol 25:149–155 doi: 10.1016/j.conb.2014.01.005.

Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes. Nat Rev Neurosci 13:407–420 doi: 10.1038/nrn3241.

Carandini M, Ferster D (2000) Membrane potential and firing rate in cat primary visual cortex. J Neurosci 20:470–484 doi: 10.1098/rspb.1986.0060.

Chialvo DR (2010) Emergent complex neural dynamics. Nat Phys 6:744–750 doi: 10.1038/nphys1803.

Churchland MM et al. (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nat Neurosci 13:369–378 doi: 10.1038/nn.2501.

Clauset A, Shalizi CR, Newman MEJ (2009) Power-Law Distributions in Empirical Data. SIAM Rev 51:661–703 doi: 10.1137/070710111.

Clawson WP, Wright NC, Wessel R, Shew WL (2017) Adaptation towards scale-free dynamics improves cortical stimulus discrimination at the cost of reduced detection. PLoS Comput Biol 13:e1005574 doi: 10.1371/journal.pcbi.1005574.

Cocchi L, Gollo LL, Zalesky A, Breakspear M (2017) Criticality in the brain: A synthesis of neurobiology, models and cognition. Prog Neurobiol 158:132–152 doi: 10.1016/j.pneurobio.2017.07.002.

Cowan JD, Neuman J, Kiewiet B, Van Drongelen W (2013) Self-organized criticality in a network of interacting neurons. J Stat Mech Theory Exp 2013:P04030 doi: 10.1088/1742-5468/2013/04/P04030.

Crochet S, Poulet JFA, Kremer Y, Petersen CCH (2011) Synaptic mechanisms underlying sparse coding of active touch. Neuron 69:1160–1175 doi: 10.1016/j.neuron.2011.02.022.

Crockett T, Wright N, Thornquist S, Ariel M, Wessel R (2015) Turtle dorsal cortex pyramidal neurons comprise two distinct cell types with indistinguishable visual responses. PLoS One 10:e0144012 doi: 10.1371/journal.pone.0144012.

Del Papa B, Priesemann V, Triesch J (2017) Criticality meets learning: Criticality signatures in a self-organizing recurrent neural network. PLoS One 12:e0178683 doi: 10.1371/journal.pone.0178683.

Deluca A, Corral Á (2013) Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions. Acta Geophys 61:1351–1394 doi: 10.2478/s11600-013-0154-9.

Denève S, Machens CK (2016) Efficient codes and balanced networks. Nat Neurosci 19:375–382 doi: 10.1038/nn.4243.

Destexhe A, Paré D (1999) Impact of Network Activity on the Integrative Properties of Neocortical Pyramidal Neurons In Vivo. J Neurophysiol 81:1531–1547 doi: 10.1152/jn.1999.81.4.1531.

Destexhe A, Rudolph M, Paré D (2003) The high-conductance state of neocortical neurons in vivo. Nat Rev Neurosci 4:739–751 doi: 10.1038/nrn1198.

Einevoll GT, Kayser C, Logothetis NK, Panzeri S (2013) Modelling and analysis of local field potentials for studying the function of cortical circuits. Nat Rev Neurosci 14:770–785 doi: 10.1038/nrn3599.

El Boustani S, Marre O, Béhuret S, Baudot P, Yger P, Bal T, Destexhe A, Frégnac Y (2009) Network-state modulation of power-law frequency-scaling in visual cortical neurons. PLoS Comput Biol 5:e1000519 doi: 10.1371/journal.pcbi.1000519.

Fingelkurts AA, Fingelkurts AA, Neves CFH (2013) Consciousness as a phenomenon in the operational architectonics of brain organization: Criticality and self-organization considerations. Chaos, Solitons & Fractals 55:13–31 doi: 10.1016/j.chaos.2013.02.007.

Freeman WJ, Cao TY (2008) Proposed Renormalization Group Analysis of Nonlinear Brain Dynamics at Criticality. In: Advances in cognitive neurodynamics : proceedings of the International Conference on Cognitive Neurodynamics, 2007 (Wang R, Shen E, Gu F, eds), pp 145–156. Springer. ISBN: 978-1-4020-8386-0.

Friedman N, Ito S, Brinkman BAW, Shimono M, Deville REL, Dahmen KA, Beggs JM, Butler TC (2012) Universal critical dynamics in high resolution neuronal avalanche data. Phys Rev Lett 108:208102 doi: 10.1103/PhysRevLett.108.208102.

Friston K (2010) The free-energy principle: A unified brain theory? Nat Rev Neurosci 11:127–138 doi: 10.1038/nrn2787.

Gal A, Marom S (2013) Self-organized criticality in single-neuron excitability. Phys Rev E - Stat Nonlinear, Soft Matter Phys 88 doi: 10.1103/PhysRevE.88.062717.

Gao L, Kostlan K, Wang Y, Wang X (2016) Distinct Subthreshold Mechanisms Underlying Rate-Coding Principles in Primate Auditory Cortex. Neuron 91:905–919 doi: 10.1016/j.neuron.2016.07.004.

Gasparini S, Magee JC (2006) State-Dependent Dendritic Computation in Hippocampal CA1 Pyramidal Neurons. J Neurosci 26:2088–2100 doi: 10.1523/JNEUROSCI.4428-05.2006.

Gatys LA, Ecker AS, Tchumatchenko T, Bethge M (2015) Synaptic unreliability facilitates information transmission in balanced cortical populations. Phys Rev E - Stat Nonlinear, Soft Matter Phys 91:62707 doi: 10.1103/PhysRevE.91.062707.

Gautam SH, Hoang TT, McClanahan K, Grady SK, Shew WL (2015) Maximizing Sensory Dynamic Range by Tuning the Cortical State to Criticality. PLoS Comput Biol 11:e1004576 doi: 10.1371/journal.pcbi.1004576.

Gidas B (1989) A Renormalization Group Approach to Image Processing Problems. IEEE Trans Pattern Anal Mach Intell 11:164–180 doi: 10.1109/34.16712.

Graupner M, Reyes AD (2013) Synaptic Input Correlations Leading to Membrane Potential Decorrelation of Spontaneous Activity in Cortex. J Neurosci 33:15075–15085 doi: 10.1523/JNEUROSCI.0347-13.2013.

Hahn G, Petermann T, Havenith MN, Yu S, Singer W, Plenz D, Nikolic D (2010) Neuronal Avalanches in Spontaneous Activity In Vivo. J Neurophysiol 104:3312–3322 doi: 10.1152/jn.00953.2009.

Haldeman C, Beggs JM (2005) Critical branching captures activity in living neural networks and maximizes the number of metastable states. Phys Rev Lett 94:58101 doi: 10.1103/PhysRevLett.94.058101.

Hammond F, Malec JF, Nick T, Buschbacher RM (2015) Part II: Statistics, Introduction. In: Handbook for Clinical Research: Design, Statistics, and Implementation, pp 77–78. New York, NY: Demos Medical. ISBN: 978-1-936287-54-3.

Hartley C, Taylor TJ, Kiss IZ, Farmer SF, Berthouze L (2014) Identification of criticality in neuronal avalanches: II. A theoretical and empirical investigation of the driven case. J Math Neurosci 4:1–42 doi: 10.1186/2190-8567-4-9.

Heeger DJ (2017) Theory of cortical function. Proc Natl Acad Sci USA 114:1773–1782 doi: 10.1073/pnas.1619788114.

Hellwig B (2000) A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. Biol Cybern 82:111–121 doi: 10.1007/PL00007964.

Helmstaedter M (2013) Cellular-resolution connectomics: Challenges of dense neural circuit reconstruction. Nat Methods 10:501–507 doi: 10.1038/nmeth.2476.

Hesse J, Gross T (2014) Self-organized criticality as a fundamental property of neural systems. Front Syst Neurosci 8:166 doi: 10.3389/fnsys.2014.00166.

Hulse BK, Lubenov E V., Siapas AG (2017) Brain State Dependence of Hippocampal Subthreshold Activity in Awake Mice. Cell Rep 18:136–147 doi: 10.1016/j.celrep.2016.11.084.

Isaacson JS, Scanziani M (2011) How inhibition shapes cortical activity. Neuron 72:231–243 doi: 10.1016/j.neuron.2011.09.027.

Jeżewski W (2004) Scaling in weighted networks and complex systems. Phys A Stat Mech its Appl 337:336–356 doi: 10.1016/j.physa.2004.01.028.

Kajikawa Y, Schroeder CE (2011) How local is the local field potential? Neuron 72:847–858 doi: 10.1016/j.neuron.2011.09.029.

Karimipanah Y, Ma Z, Miller JK, Yuste R, Wessel R (2017a) Neocortical activity is stimulus- and scale-invariant. PLoS One 12:e0177396 doi: 10.1371/journal.pone.0177396.

Karimipanah Y, Ma Z, Wessel R (2017b) Criticality predicts maximum irregularity in recurrent networks of excitatory nodes. PLoS One 12:e0182501 doi: 10.1371/journal.pone.0182501.

Kello CT (2013) Critical branching neural networks. Psychol Rev 120:230–254 doi: 10.1037/a0030970.

Kelly RC, Smith MA, Kass RE, Lee TS (2010) Local field potentials indicate network state and account for neuronal response variability. J Comput Neurosci 29:567–579 doi: 10.1007/s10827-009-0208-9.

Kerby DS (2014) The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. Compr Psychol 3:11.IT.3.1 doi: 10.2466/11.IT.3.1.

Kinouchi O, Copelli M (2006) Optimal dynamical range of excitable networks at criticality. Nat Phys 2:348–351 doi: 10.1038/nphys289.

Klaus A, Yu S, Plenz D (2011) Statistical analyses support power law distributions found in neuronal avalanches. PLoS One 6:e19779 doi: 10.1371/journal.pone.0019779.

Koch-Janusz M, Ringel Z (2018) Mutual information, neural networks and the renormalization group. Nat Phys 14:578–582 doi: 10.1038/s41567-018-0081-4.

Kodandaramaiah SB, Franzesi GT, Chow BY, Boyden ES, Forest CR (2012) Automated whole-cell patch-clamp electrophysiology of neurons in vivo. Nat Methods 9:585–587 doi: 10.1038/nmeth.1993.

Larkum ME, Watanabe S, Lasser-Ross N, Rhodes P, Ross WN (2008) Dendritic Properties of Turtle Pyramidal Neurons. J Neurophysiol 99:683–694 doi: 10.1152/jn.01076.2007.

Larremore DB, Carpenter MY, Ott E, Restrepo JG (2012) Statistical properties of avalanches in networks. Phys Rev E - Stat Nonlinear, Soft Matter Phys 85:66131 doi: 10.1103/PhysRevE.85.066131.

Larremore DB, Shew WL, Ott E, Restrepo JG (2011a) Effects of network topology, transmission delays, and refractoriness on the response of coupled excitable systems to a stochastic stimulus. Chaos 21:25117 doi: 10.1063/1.3600760.

Larremore DB, Shew WL, Ott E, Sorrentino F, Restrepo JG (2014) Inhibition causes ceaseless dynamics in networks of excitable nodes. Phys Rev Lett 112:138103 doi: 10.1103/PhysRevLett.112.138103.

Larremore DB, Shew WL, Restrepo JG (2011b) Predicting criticality and dynamic range in complex networks: Effects of topology. Phys Rev Lett 106:58101 doi: 10.1103/PhysRevLett.106.058101.

Laughlin RB (2014) Physics, Emergence, and the Connectome. Neuron 83:1253–1255 doi: 10.1016/j.neuron.2014.08.006.

Le Doussal P, Wiese KJ (2009) Size distributions of shocks and static avalanches from the functional renormalization group. Phys Rev E - Stat Nonlinear, Soft Matter Phys 79:51106 doi: 10.1016/j.amjmed.2015.02.003.

Lee AK, Brecht M (2018) Elucidating Neuronal Mechanisms Using Intracellular Recordings during Behavior. Trends Neurosci 41:385–403 doi: 10.1016/j.tins.2018.03.014.

Lee WCA, Bonin V, Reed M, Graham BJ, Hood G, Glattfelder K, Reid RC (2016) Anatomy and function of an excitatory network in the visual cortex. Nature 532:370–374 doi: 10.1038/nature17192.

Lee YC, Doolen G, Chen HH, Sun GZ, Maxwell T, Lee HY, Giles CL (1986) Machine Learning Using a Higher Order Correlation Network. Phys D 22:276–306 doi: 10.1016/0167-2789(86)90300-6.

Levina A, Herrmann JM, Denker M (2007) Critical branching processes in neural networks. Pamm 7:1030701–1030702 doi: 10.1002/pamm.200700029.

Levina A, Priesemann V (2017) Subsampling scaling. Nat Commun 8:15140 doi: 10.1038/ncomms15140.

Litwin-Kumar A, Doiron B (2012) Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat Neurosci 15:1498–1505 doi: 10.1038/nn.3220.

Liu J, Newsome WT (2006) Local Field Potential in Cortical Area MT: Stimulus Tuning and Behavioral Correlations. J Neurosci 26:7779–7790 doi: 10.1523/JNEUROSCI.5052-05.2006.

London M, Häusser M (2005) Dendritic Computation. Annu Rev Neurosci 28:503–532 doi: 10.1146/annurev.neuro.28.061604.135703.

Magee JC (2000) Dendritic integration of excitatory synaptic input. Nat Rev Neurosci 1:181–190 doi: 10.1038/35044552.

Malina KCK, Mohar B, Rappaport AN, Lampl I (2016) Local and thalamic origins of correlated ongoing and sensory-evoked cortical activities. Nat Commun 7:12740 doi: 10.1038/ncomms12740.

Mao BQ, Hamzei-Sichani F, Aronov D, Froemke RC, Yuste R (2001) Dynamics of spontaneous activity in neocortical slices. Neuron 32:883–898 doi: 10.1016/S0896-6273(01)00518-9.

Maris HJ, Kadanoff LP (1978) Teaching the renormalization group. Am J Phys 46:652–657 doi: 10.1119/1.11224.

Marshall N, Timme NM, Bennett N, Ripp M, Lautzenhiser E, Beggs JM (2016) Analysis of Power Laws, Shape Collapses, and Neural Complexity: New Techniques and MATLAB Support via the NCC Toolbox. Front Physiol 7:250 doi: 10.1109/EuCAP.2016.7481967.

Mastrogiuseppe F, Ostojic S (2018) Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks. Neuron 99:609–623.e29 doi: 10.1016/j.neuron.2018.07.003.

McGinley MJ, David S V., McCormick DA (2015) Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. Neuron 87:179–192 doi: 10.1016/j.neuron.2015.05.038.

Mehta P, Schwab DJ (2014) An exact mapping between the Variational Renormalization Group and Deep Learning. arXiv:14103831 doi: 10.1093/mnras/stv632.

Meinecke DL, Peters A (1987) GABA immunoreactive neurons in rat visual cortex. J Comp Neurol 261:388–404 doi: 10.1002/cne.902610305.

Millman D, Mihalas S, Kirkwood A, Niebur E (2010) Self-organized criticality occurs in non-conservative neuronal networks during “up” states. Nat Phys 6:801–805 doi: 10.1038/nphys1757.

Mokeichev A, Okun M, Barak O, Katz Y, Ben-Shahar O, Lampl I (2007) Stochastic Emergence of Repeating Cortical Motifs in Spontaneous Membrane Potential Fluctuations In Vivo. Neuron 53:413–425 doi: 10.1016/j.neuron.2007.01.017.

Moore JJ, Ravassard PM, Ho D, Acharya L, Kees AL, Vuong C, Mehta MR (2017) Dynamics of cortical dendritic membrane potential and spikes in freely behaving rats. Science (80-) 355:eaaj1497 doi: 10.1126/science.aaj1497.

Moretti P, Munoz MA (2013) Griffiths phases and the stretching of criticality in brain networks. Nat Commun 4:2521 doi: 10.1038/ncomms3521.

Mulligan KA, Ulinski PS (1990) Organization of geniculocortical projections in turtles: Isoazimuth lamellae in the visual cortex. J Comp Neurol 296:531–547 doi: 10.1002/cne.902960403.

Ness T V., Remme MWH, Einevoll GT (2016) Active subthreshold dendritic conductances shape the local field potential. J Physiol 594:3809–3825 doi: 10.1113/JP272022.

Nishimori H, Ortiz G (2011) Elements of Phase Transitions and Critical Phenomena. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780199577224.001.0001.

Nunez PL, Srinivasan R, Ingber L (2013) Theoretical and Experimental Electrophysiology in Human Neocortex: Multiscale Dynamic Correlates of Conscious Experience. In: Multiscale Analysis and Nonlinear Dynamics (Pesenson M (Meyer) Z, ed), pp 147–177. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. ISBN: 9783527671632.

Okun M, Steinmetz NA, Cossell L, Iacaruso MF, Ko H, Barthó P, Moore T, Hofer SB, Mrsic-Flogel TD, Carandini M, Harris KD (2015) Diverse coupling of neurons to populations in sensory cortex. Nature 521:511–515 doi: 10.1038/nature14273.

Papanikolaou S, Bohn F, Sommer RL, Durin G, Zapperi S, Sethna JP (2011) Universality beyond power laws and the average avalanche shape. Nat Phys 7:316–320 doi: 10.1038/nphys1884.

Perin R, Berger TK, Markram H (2011) A synaptic organizing principle for cortical neuronal groups. Proc Natl Acad Sci 108:5419–5424 doi: 10.1073/pnas.1016051108.

Petersen CCH (2017) Whole-Cell Recording of Neuronal Membrane Potential during Behavior. Neuron 95:1266–1281 doi: 10.1016/j.neuron.2017.06.049.

Pettersen KH, Lindén H, Tetzlaff T, Einevoll GT (2014) Power Laws from Linear Neuronal Cable Theory: Power Spectral Densities of the Soma Potential, Soma Membrane Current and Single-Neuron Contribution to the EEG. PLoS Comput Biol 10:e1003928 doi: 10.1371/journal.pcbi.1003928.

Poil S-S, Hardstone R, Mansvelder HD, Linkenkaer-Hansen K (2012) Critical-State Dynamics of Avalanches and Oscillations Jointly Emerge from Balanced Excitation/Inhibition in Neuronal Networks. J Neurosci 32:9817–9823 doi: 10.1523/JNEUROSCI.5990-11.2012.

Poil SS, Van Ooyen A, Linkenkaer-Hansen K (2008) Avalanche dynamics of human brain oscillations: Relation to critical branching processes and temporal correlations. Hum Brain Mapp 29:770–777 doi: 10.1002/hbm.20590.

Porta LD, Copelli M (2018) Modeling neuronal avalanches and long-range temporal correlations at the emergence of collective oscillations: continuously varying exponents mimic M/EEG results. bioRxiv:423921 doi: 10.1101/423921.

Poulet JFA, Petersen CCH (2008) Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. Nature 454:881–885 doi: 10.1038/nature07150.

Priesemann V (2014) Spike avalanches in vivo suggest a driven, slightly subcritical brain state. Front Syst Neurosci 8:108 doi: 10.3389/fnsys.2014.00108.

Priesemann V, Munk MHJ, Wibral M (2009) Subsampling effects in neuronal avalanche distributions recorded in vivo. BMC Neurosci 10:40 doi: 10.1186/1471-2202-10-40.

Priesemann V, Shriki O (2018) Can a time varying external drive give rise to apparent criticality in neural systems? PLoS Comput Biol 14:e1006081 doi: 10.1371/journal.pcbi.1006081.

Pruessner G (2012) Self-Organized Criticality Theory Models and Characterisation. Cambridge, UK ; New York: Cambridge University Press. ISBN: 9780521853354.

Rapp PE, Albano AM, Zimmerman ID, Jiménez-Montaño MA (1994) Phase-randomized surrogates can produce spurious identifications of non-random structure. Phys Lett A 192:27–33 doi: 10.1016/0375-9601(94)91010-3.

Ribeiro TL, Copelli M, Caixeta F, Belchior H, Chialvo DR, Nicolelis MAL, Ribeiro S (2010) Spike avalanches exhibit universal dynamics across the sleep-wake cycle. PLoS One 5:e14129 doi: 10.1371/journal.pone.0014129.

Ribeiro TL, Ribeiro S, Belchior H, Caixeta F, Copelli M (2014) Undersampled critical branching processes on small-world and random networks fail to reproduce the statistics of spike avalanches. PLoS One 9:e94992 doi: 10.1371/journal.pone.0094992.

Roberts S (2018) Mathematician Carina Curto Thinks Like a Physicist to Solve Neuroscience Problems. Quanta Mag.

Sachdev RNS, Ebner FF, Wilson CJ (2004) Effect of Subthreshold Up and Down States on the Whisker-Evoked Response in Somatosensory Cortex. J Neurophysiol 92:3511–3521 doi: 10.1152/jn.00347.2004.

Sachidhanandam S, Sreenivasan V, Kyriakatos A, Kremer Y, Petersen CCH (2013) Membrane potential correlates of sensory perception in mouse barrel cortex. Nat Neurosci 16:1671–1677 doi: 10.1038/nn.3532.

Saha D, Morton D, Ariel M, Wessel R (2011) Response properties of visual neurons in the turtle nucleus isthmi. J Comp Physiol A Neuroethol Sensory, Neural, Behav Physiol 197:153–165 doi: 10.1007/s00359-010-0596-3.

Saremi S, Sejnowski TJ (2016) Correlated Percolation, Fractal Structures, and Scale-Invariant Distribution of Clusters in Natural Images. IEEE Trans Pattern Anal Mach Intell 38:1016–1020 doi: 10.1109/TPAMI.2015.2481402.

Scarpetta S, Apicella I, Minati L, De Candia A (2018) Hysteresis, neural avalanches, and critical behavior near a first-order transition of a spiking neural network. Phys Rev E 97:62305 doi: 10.1103/PhysRevE.97.062305.

Scarpetta S, Candia A de (2013) Neural Avalanches at the Critical Point between Replay and Non-Replay of Spatiotemporal Patterns. PLoS One 8:e64162 doi: 10.1371/journal.pone.0064162.

Schwab DJ, Nemenman I, Mehta P (2014) Zipf's law and criticality in multivariate data without fine-tuning. Phys Rev Lett 113:68102 doi: 10.1103/PhysRevLett.113.068102.

Senseman DM, Robbins KA (1999) Modal behavior of cortical neural networks during visual processing. J Neurosci 19:RC3 (1-7) doi: 10.1523/JNEUROSCI.19-10-j0004.1999.

Senseman DM, Robbins KA (2002) High-Speed VSD Imaging of Visually Evoked Cortical Waves: Decomposition Into Intra- and Intercortical Wave Motions. J Neurophysiol 87:1499–1514 doi: 10.1152/jn.00475.2001.

Sethna JP, Dahmen KA, Myers CR (2001) Crackling noise. Nature 410:242–250 doi: 10.1038/35065675.

Sfondrini A (2012) Introduction to universality and renormalization group techniques. Proc Sci 195 ISBN: 9783540646662.

Shaukat A, Thivierge J-P (2016) Statistical Evaluation of Waveform Collapse Reveals Scale-Free Properties of Neuronal Avalanches. Front Comput Neurosci 10:29 doi: 10.3389/fncom.2016.00029.

Shew WL, Clawson WP, Pobst J, Karimipanah Y, Wright NC, Wessel R (2015) Adaptation to sensory input tunes visual cortex to criticality. Nat Phys 11:659–663 doi: 10.1038/nphys3370.

Shew WL, Plenz D (2012) The Functional Benefits of Criticality in the Cortex. Neuroscientist 19:88–100 doi: 10.1177/1073858412445487.

Shew WL, Yang H, Yu S, Roy R, Plenz D (2011) Information Capacity and Transmission Are Maximized in Balanced Cortical Networks with Neuronal Avalanches. J Neurosci 31:55–63 doi: 10.1523/JNEUROSCI.4637-10.2011.

Shimono M, Beggs JM (2015) Functional clusters, hubs, and communities in the cortical microconnectome. Cereb Cortex 25:3743–3757 doi: 10.1093/cercor/bhu252.

Shoham S, O'Connor DH, Segev R (2006) How silent is the brain: Is there a "dark matter" problem in neuroscience? J Comp Physiol A Neuroethol Sensory, Neural, Behav Physiol 192:777–784 doi: 10.1007/s00359-006-0117-6.

Shriki O, Alstott J, Carver F, Holroyd T, Henson RNA, Smith ML, Coppola R, Bullmore E, Plenz D (2013) Neuronal Avalanches in the Resting MEG of the Human Brain. J Neurosci 33:7079–7090 doi: 10.1523/JNEUROSCI.4286-12.2013.

Shriki O, Yellin D (2016) Optimal Information Representation and Criticality in an Adaptive Sensory Recurrent Neuronal Network. PLoS Comput Biol 12:e1004698 doi: 10.1371/journal.pcbi.1004698.

Song S, Reigl M, Nelson S, Chklovskii DB, Sjo PJ (2005) Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits. PLoS Biol 3 doi: 10.1371/journal.pbio.0030068.

Stanley HE (1999) Scaling, universality, and renormalization: Three pillars of modern critical phenomena. Rev Mod Phys 71:S358–S366 doi: 10.1103/RevModPhys.71.S358.

Stepanyants A, Hof PR, Chklovskii DB (2002) Geometry and structural plasticity of synaptic connectivity. Neuron 34:275–288 doi: 10.1016/S0896-6273(02)00652-9.

Stuart GJ, Spruston N (2015) Dendritic integration: 60 years of progress. Nat Neurosci 18:1713–1721 doi: 10.1038/nn.4157.

Taillefumier T, Magnasco MO (2013) A phase transition in the first passage of a Brownian process through a fluctuating boundary with implications for neural coding. Proc Natl Acad Sci 110:E1438–E1443 doi: 10.1073/pnas.1212479110.

Tang C, Bak P (1988) Critical exponents and scaling relations for self-organized critical phenomena. Phys Rev Lett 60:2347–2350 doi: 10.1103/PhysRevLett.60.2347.

Taylor TJ, Hartley C, Simon PL, Kiss IZ, Berthouze L (2013) Identification of criticality in neuronal avalanches: I. A theoretical investigation of the non-driven case. J Math Neurosci 3:1–26 doi: 10.1186/2190-8567-3-51-26.

Theiler J, Eubank S, Longtin A, Galdrikian B, Doyne Farmer J (1992) Testing for nonlinearity in time series: the method of surrogate data. Phys D Nonlinear Phenom 58:77–94 doi: 10.1016/0167-2789(92)90102-S.

Timme NM, Marshall NJ, Bennett N, Ripp M, Lautzenhiser E, Beggs JM (2016) Criticality maximizes complexity in neural tissue. Front Physiol 7:425 doi: 10.3389/fphys.2016.00425.

Tishby N, Zaslavsky N (2015) Deep learning and the information bottleneck principle. In: 2015 IEEE Information Theory Workshop (ITW), pp 1–5. Jerusalem. doi: 10.1109/ITW.2015.7133169.

Tomen N, Rotermund D, Ernst U (2014) Marginally subcritical dynamics explain enhanced stimulus discriminability under attention. Front Syst Neurosci 8:151 doi: 10.3389/fnsys.2014.00151.

Touboul J, Destexhe A (2010) Can power-law scaling and neuronal avalanches arise from stochastic dynamics? PLoS One 5:e8982 doi: 10.1371/journal.pone.0008982.

Touboul J, Destexhe A (2017) Power-law statistics and universal scaling in the absence of criticality. Phys Rev E 95:12413 doi: 10.1103/PhysRevE.95.012413.

Tsodyks M V., Markram H (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proc Natl Acad Sci 94:719–723 doi: 10.1073/pnas.94.2.719.

Ulinski PS (1990) The Cerebral Cortex of Reptiles. In: Comparative Structure and Evolution of Cerebral Cortex, Part I (Jones EG, Peters A, eds), pp 139–215. New York: Plenum Press. doi: 10.1007/978-1-4757-9622-3_5.

Werner G (2012) From brain states to mental phenomena via phase space transitions and renormalization group transformation: proposal of a theory. Cogn Neurodyn 6:199–202 doi: 10.1007/s11571-011-9187-4.

Wertz A, Trenholm S, Yonehara K, Hillier D, Raics Z, Leinweber M, Szalay G, Ghanem A, Keller G, Rózsa B, Conzelmann KK, Roska B (2015) Single-cell-initiated monosynaptic tracing reveals layer-specific cortical network modules. Science (80-) 349:70–74 doi: 10.1126/science.aab1687.

Willcox CR (1991) Understanding hierarchical neural network behaviour: a renormalization group approach. J Phys A Math Theor 24:2655 doi: 10.1088/0305-4470/24/11/030.

Williams-García R V., Moore M, Beggs JM, Ortiz G (2014) Quasicritical brain dynamics on a nonequilibrium Widom line. Phys Rev E - Stat Nonlinear, Soft Matter Phys 90:62714 doi: 10.1103/PhysRevE.90.062714.

Wilson DE, Whitney DE, Scholl B, Fitzpatrick D (2016) Orientation selectivity and the functional clustering of synaptic inputs in primary visual cortex. Nat Neurosci 19:1003–1009 doi: 10.1038/nn.4323.

Wilting J, Priesemann V (2018) Inferring collective dynamical states from widely unobserved systems. Nat Commun 9 doi: 10.1038/s41467-018-04725-4.

Wolfe J, Houweling AR, Brecht M (2010) Sparse and powerful cortical spikes. Curr Opin Neurobiol 20:306–312 doi: 10.1016/j.conb.2010.03.006.

Wright NC, Hoseini MS, Wessel R (2017a) Adaptation modulates correlated subthreshold response variability in visual cortex. J Neurophysiol 118:1257–1269 doi: 10.1152/jn.00124.2017.

Wright NC, Hoseini MS, Yasar TB, Wessel R (2017b) Coupling of synaptic inputs to local cortical activity differs among neurons and adapts after stimulus onset. J Neurophysiol 118:3345–3359 doi: 10.1152/jn.00398.2017.

Wright NC, Wessel R (2017) Network activity influences the subthreshold and spiking visual responses of pyramidal neurons in the three-layer turtle cortex. J Neurophysiol 118:jn.00340.2017 doi: 10.1152/jn.00340.2017.

Yaşar TB, Wright NC, Wessel R (2016) Inferring presynaptic population spiking from single-trial membrane potential recordings. J Neurosci Methods 259:13–21 doi: 10.1016/j.jneumeth.2015.11.019.

Yu S, Klaus A, Yang H, Plenz D (2014) Scale-invariant neuronal avalanche dynamics and the cut-off in size distributions. PLoS One 9:e99761 doi: 10.1371/journal.pone.0099761.

Chapter 3:
Stimulus tuned dynamical trajectories underlie synaptically driven transmembrane signals in visual cortex

Cortical neurons are like embedded sensors polling a subnetwork of neurons which spans brain regions. Often such neural populations are characterized with non-linear dynamical systems paradigms, such as attractor networks. The prospect of obtaining population dynamics from a naturally in-situ and fully integrated informant (a single neuron) motivates a machine learning algorithm to find attractor-like dynamical rules in synaptically driven fluctuations of transmembrane currents and potential. Since the population dynamics we seek is largely driven by visual responses, we can validate the algorithm by finding stimulus information not evinced by existing methods that average out fluctuations. We obtained intracellular recordings of fluctuating transmembrane current and potential in visual cortex (V1, L2/3) of mice in response to visual stimulation with drifting gratings. We estimated equations governing 3D trajectories produced by time-delay dimensionality expansion of these recordings. Fine distinctions about orientations of drifting gratings existed in the equations governing trajectories from synaptic excitation, a finding undiscoverable using older methods. Broadly, governing equations contained more fine-grained information about size and contrast than firing rate substitutes and evinced distinct regimes of presynaptic population dynamics.

3.1 Introduction:

Pyramidal neurons aggregate population synaptic transmissions and neuromodulatory information then pass limited but useful information to downstream neuronal populations via action potentials. Experimentally accessing such upstream information would allow researchers to analyze the behavior of an intrinsically unambiguous neuroanatomical subpopulation: the group of neurons that synapse onto the same neuron (or neurons) (Yaşar et al., 2016). Spurts of spiking among interacting populations of upstream neurons cause spurts of synaptic input. Synaptic inputs cause propagating pulses that interfere with each other and are further transformed by single-neuron biophysical effects, including membrane properties (London and Häusser, 2005; Yaşar et al., 2016; Wright et al., 2017). These actions throughout dendritic structures result in fluctuations of membrane current and membrane potential at the neuronal cell body, i.e., the soma.

However, network state signatures can be uncovered with long enough recordings (Johnson et al., 2019). Thanks to the ongoing revolution in data analysis (Bzdok and Yeo, 2017) and neuronal recording methods (Kim et al., 2017; Jouhanneau and Poulet, 2019) information focused on in-situ presynaptic populations would be uniquely insightful. Electrode array or imaging methods record an amalgam of neuron types and may span groups with different functions (e.g. tunings). Neuronal population dynamics are often characterized by systems of differential equations (Rabinovich et al., 2006; Pandarinath et al., 2018). Therefore, time series analysis and latent variable discovery may be the missing

elements needed for extracting detailed population information within single trials of single-neuron intracellular recordings.

A dynamical systems perspective provides a hypothesis for how to extract population information from postsynaptic fluctuations. The attractor network paradigm (Amit and Amit, 1992; Eliasmith, 2005; Wu and Amari, 2005; Chambers and Rumpel, 2017), explains representation by stochastic spatiotemporal patterns of spiking through modelling neural networks as very high dimensional dynamical systems. This paradigm is novel in primary visual cortex (Goldberg et al., 2004; Miconi et al., 2016). Attractor network dynamics exhibit attracting sets representing brain operations, i.e. a stimulus evokes a perturbation causing network dynamics to either explore state space near a different fixed point, or to undergo a bifurcation. Although formal governing equations have fixed parameters, approximations and reductions have different parameters near different fixed points. Thus, stimuli should modulate the parameters of rudimentary equations fitted to brief snippets of evoked activity. This modulation is illustrated in Figure 3.1a by pendulum with physical properties which are like stimulus characteristics providing the context for a recording of neural activity. Upon change of coordinates, the pendulum's motion is a simple Lorenz system (Clerc et al., 2001) (Figure 3.1b) with coefficients modulated by physical properties (i.e. context).

To test our dynamical systems insights we recorded membrane current and potential from single neurons in primary visual cortex of awake mice concurrent with visual stimulation by drifting gratings of varying orientation, size, and contrast (Figure 3.1c). We applied a

Whitney-Takens time-delay based dimensionality expansion (Lainscsek and Sejnowski, 2015; Oprisan et al., 2015) to one-dimensional intracellular recordings. Thus obtaining moderate dimension neural trajectories that represented a projection of the putative high-dimensional dynamical system onto this lower dimensional space. We then applied a model discovery algorithm (Brunton et al., 2016) (Figure 3.1d) modified to interpret the coefficients of differential equations as a basis for stimulus representation. We call the algorithm "dynamical discrimination".

We found that dimensionality expansion of our intracellular recordings followed by dynamical discrimination permited better than chance classification of small changes in orientation, size, and contrast of drifting gratings. Correct classification rates exceeded classifiers using the magnitude of transmembrane current and potential deflection (a firing rate substitute). This is especially significant for grating orientation because neurons selectively fire for a more narrow range of orientations in V1 than in the thalamus which provides inputs to V1 (Seriès et al., 2004; Li et al., 2012; Sun et al., 2016), but deflection failed to reflect small changes in orientation. In contrast, classification on summaries of trajectories failed because trajectory derivatives encoded such stimulus details.

Ultimately, dynamical discrimination connects model discovery (Daniels and Nemenman, 2015; Brunton et al., 2016) to latent variable discovery (Gallego et al., 2018; Pandarinath et al., 2018; Whiteway and Butts, 2019), and exemplifies machine learning to test scientific hypotheses (Bishop, 2013; Butner et al., 2019) because it is premised on an attractor network hypothesis. Thus, it enables new methods of quantifying attractor network

principles in the primary visual cortex (Goldberg et al., 2004; Miconi et al., 2016) and single-neuron intracellular recordings.

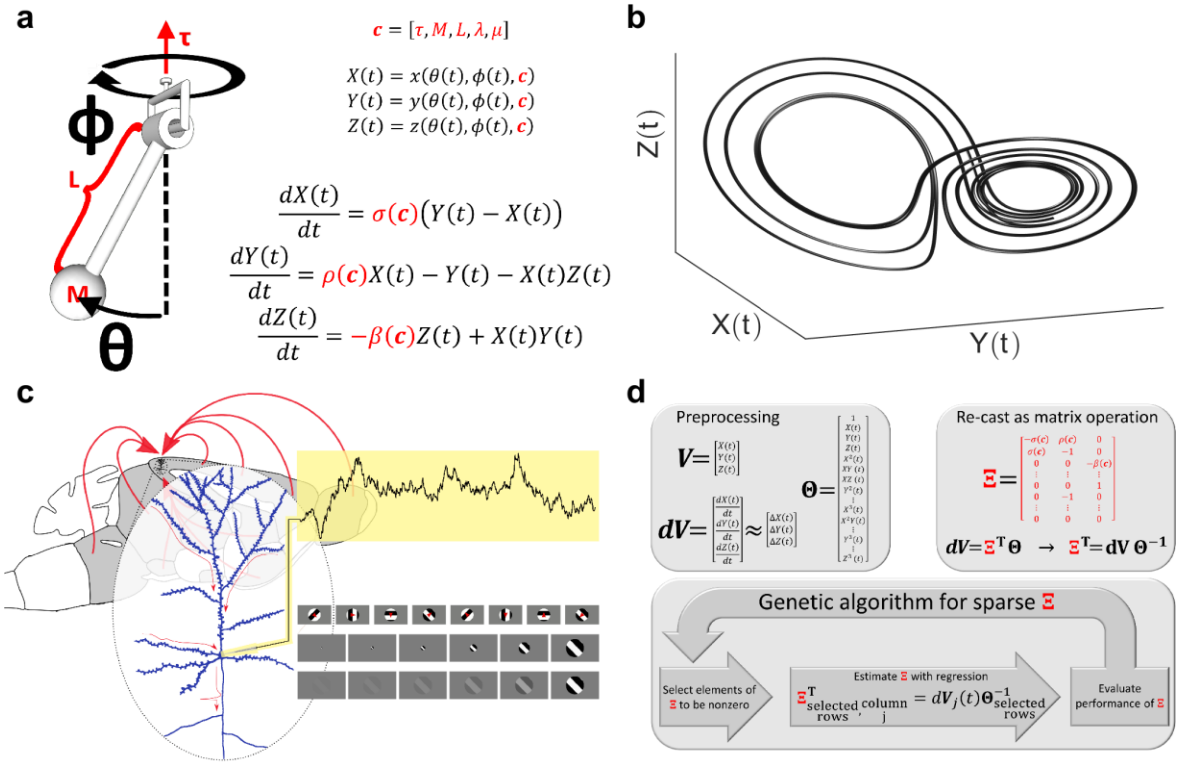

Figure 3.1 | The context of a dynamical trajectory is represented in governing equation parameters. **a**, A pendulum governed by Lorenz equations after changing variables from spherical coordinates (θ, ϕ, r) to abstract coordinates (X, Y, Z) (Clerc et al., 2001). Our interpretation of "context" is illustrated by the parameters: constant torque τ, bob mass M, rod length L, Stoke's Law coefficient λ, and friction μ. **b**, A chaotic Lorenz attractor (σ=10, β=8/3, ρ=28. **c**, An illustration of our data source: intracellular recordings from single neurons in mouse V1. The context of recorded dynamics is influenced by visual stimulation. We have three stimulus categories: eight drifting grating orientations (top), six logarithmically spaced sizes (center), and six logarithmic contrast levels (bottom). **d**, We identify context through modified Sparse Identification of Nonlinear Dynamics (SINDy). We illustrate a context modulated sparse dynamical systems coefficient matrix, Ξ, with the Lorenz system (see a). V is a time-delay embedding of a single dynamical variable (e.g. membrane potential). Θ contains all polynomial combinations of V up to cubic terms. dV is a derivative estimate. Ξ, is learned through linear regression to regress Θ onto dV. Ξ represents our data compactly enough to train classifiers to predict context (e.g. stimulus) with few examples. A genetic algorithm finds Ξ by picking different sets of nonzero elements at each generation. The best set of nonzero elements is chosen based on whether Ξ leads either to the best stimulus classification performance or best trajectory reconstruction while maintaining sparseness.

3.2 Methods and Materials

The experimental data was originally gathered to support another publication (Adesnik, 2017). The methods are covered there and relevant details are repeated here (Harriman and Patel, 2014).

All procedures were approved by the University of California, Berkeley ACUC. Wild-type (C57;B6 x ICR white), emx1-IRES-Cre, and SOM-IRES-Cre mice were used. Mice of both sexes were used equally, and no differences were observed between sexes. For *in vivo* recordings mice were 5-14 weeks old.

3.2.1 Animals: Surgery and Electrophysiological Recording

Mice were headplated under isoflurane (1.5%–2%) anesthesia with a small stainless steel plate, attached to the skull with Metabond. The skull was protected with cyanoacrylate glue and dental cement (Orthojet). 1-7 days post surgery, Mice were habituated to run freely on a small, 6”diameter rotating disc during head fixation. On the day of surgery mice were anesthetized with 1.5%–2% iso-flurane and a small craniotomy was made over V1 by removing the dental cement and slowly thinning the skull until it was trans-parent with a 0.25 mm carbide burr. A small stainless steel needle (27G) was used to open a hole 150-500 um in diameter over V1 with no or minimal bleeding. The dura was always left intact. The craniotomy was covered with sterile saline and the animal was allowed to recover under fixation for 15-30 min prior to whole-cell recording. Animals typically began running on the treadmill immediately upon arousal, and either continuously or intermittently thereafter. Under these experimental conditions mouse move their eyes only infrequently,

and most ocular deviations are too small to significantly impact neuronal responses (Adesnik et al., 2012), and the pupil was not tracked.

3.2.2 Electrophysiology

Prior to intracellular experiments, a patch pipette filled with ACSF (in mM: NaCl 119, KCl 2.5, $MgSO_4$ 1.3, NaH2PO_4 1.3, glucose 20, $NaHCO_3$ 26, $CaCl_2$ 2.5) was lowered slowly into the L2/3 under visual guidance (Leica MZ6 stereomicroscope). Using multiunit ac-tivity and the LFP as a guide, the visual receptive field of the corresponding location for subsequent whole cell recording was mapped via a hand-controlled small circle (5 degrees) of changing contrast on the visual stimulus monitor (more details below). This elec-trode was then removed, and patch pipettes were then inserted in same manner for intracellular recording containing: $CsMeSO_4$ (for voltage clamp) or KGluconate (for current clamp) 135 mM, NaCl 8 mM, HEPES 10 mM, Na_3GTP 0.3 mM, MgATP 4 mM, EGTA 0.3 mM, QX-314-Cl 5 mM (voltage clamp only), TEA-Cl 5mM (voltage clamp only). Although the cells were patched with the blind approach, the conditions used have been reported to strongly bias recording to regular-spiking putative pyramidal cells (Liu et al., 2009). Nevertheless, the data reported here is likely to come from a mix of cell types, dominated nevertheless by excitatory neurons, which make up the majority of L2/3 cells.

Under these conditions, in Vclamp, the mean series resistance, prior to any compensation, was 18 ± 1 MΩ across the recording sessions, and fairly stable (Adesnik, 2017). It is now well established that locomotion and/or brain state influence spontaneous activity and sensory responses in V1 (Niell and Stryker, 2008; Ayaz et al., 2013; Reimer et al., 2014;

Vinck et al., 2015), although the exact mechanisms underlying these changes remain a matter of debate (Polack et al., 2013; Fu et al., 2014; Pakan et al., 2016). Consistent with prior findings, during locomotion (Bennett et al., 2013), visually evoked E and I were significantly increased (E: not running: 70 ± 6 pC/s, running: 81 ± 8 pC/s, $n = 39$ cells, $p < 0.005$; I: not running: 114 ± 12 pC/s, running: 159 ± 20 pC/s, $n = 39$ cells, $p < 0$ 0.005, Wilcoxon sign rank test (Adesnik, 2017)). Conversely, spontaneous excitation and inhibition, as well as the mean input conductance in the absence of a stimulus showed no significant change (E: $p = 0.9$; I: $p = 0.4$, input resistance: $p = 0.93$, $n = 39$ cells, Wilcoxon sign rank test (Adesnik, 2017)).

Both extracellular and intracellular experiments employed an Axopatch 200B amplifier. All data were acquired with custom software written in MATLAB using a National Instruments PCIe-6353 card. Glass pipettes (Sutter instruments) containing either a potassium based internal (for measurements of membrane potential and spiking) or cesium (with added QX-314-Cl, and tetraethy-lammoniaum-Cl) for voltage clamp recording, were used. Pipettes were pulled on a Sutter P1000 puller in a two stage pull to a long taper pipette of a resistance between 3-5 MOhm. To insert the electrode into the small craniotomy, the ACSF on the skull was removed and the craniotomy briefly dried with compressed air. The electrode was mounted on a Sutter MP285 manipulator, lowered until it nearly reached the brain surface, then the chamber formed by the headplate and cement was re-filled with ACSF, all under visual guidance. The pipette resistance was checked via an oscilloscope and a constant 5 mV voltage step in voltage clamp. High positive pressure (150 mbar) was applied to the pipette, and it was lowered until a brief and rapid increase in pipette

resistance was observed, indicating contact with the dura. The pipette was zeroed to obtain an accurate measurement of recording depth, and then the pipette was advanced quickly through dura, and only pipettes that quickly returned to their baseline resistance were advanced further, otherwise they were exchanged for a fresh pipette and the process was repeated. Once inside the brain the pres-sure was quickly lowered to 10-30 mBar to search for L2/3 neurons via abrupt, 'bounce' like changes in pipette resistance indicating contact with a plasma membrane, using pulsatile steps of the manipulator (1-2 microns). Upon apparent contact, pipette pressure was released, and slight positive pressure was used to obtain a gigaohm seal. Pipette capacitance was then neutralized and the membrane ruptured by brief suction pulses. Upon rupture the whole cell access was optimized by either slow negative or (more typi-cally) positive pressure and locked off. In the first 2-4 min the receptive field of the cell (either via membrane potential, spiking, or excitatory current, command potential =-70 mV) was remapped in the same manner as above, to center the stimulus on the re-corded cell's receptive field (almost always aligned with the previous measurement from extracellular recording). The orientation of the stimulus was also optimized for each cell. After spontaneous and evoked responses stabilized (typically 2-4 min) experiments were commenced. Membrane potential was obtained in voltage following mode (current clamp) with no current injection. For voltage clamped cells, cells were clamped either at -70 mV to measure synaptic excitation (approximate reversal potential for inhibition), or at +10 mV to measure synaptic inhibition (approximate reversal potential for excitation), uncorrected for the junction potential. Series resistance was monitored on every trial with a negative voltage step. Cells were only included if their

series resistance stayed within 20% of their initial value, passively or by adjusting pipette pressure.

3.2.3 Visual Stimulation

Visual stimuli were generated with Psychophysics toolbox (Brainard, 1997) using custom software in MATLAB (MathWorks) and pre-sented on a gamma corrected 23-inch Eizo FORIS FS2333 LCD display with a 60-Hz refresh rate. Stimuli consisted of drifting square wave gratings with contrast, size, or orientation varied, while all other parameters remained fixed, at 0.04 cycles per degree and 2-2.5 cycles per second. In experiments with varying contrast, size was fixed at 12 degrees, and the orientation fixed at the preferred orien-tation of the cell (measured via spike rate, V_m depolarization, or mean synaptic excitation). In 7/12 cells contrast was varied in six log increments from 1%–100%, and in 5/12 cells from 10%–100%. In experiments with varying size, contrast was set at 100% at the orientation set as above. The grating drifted immediately upon display, and lasted 0.6-1.5 s. Inter-trial-intervals (gray screen) lasted from 1.5-3 s.

3.2.4 Tests and measures

For making claims about whether one algorithm performed significantly better than another set we used the Wilcoxon signed-rank (Wilcoxon, 1992) method implemented with MATLAB. This tests whether the median difference between two matched sets is significant. When comparing whether synaptic excitation or inhibition performed better for the same algorithm while limiting the comparison to just one stimulus variable

(orientation, size, contrast) we used the Wilcoxon signed-rank test because each cell was recorded from in both modes. If we were making the same comparison across multiple stimulus variables, or comparing membrane potential to other recording modes we used the Wilcoxon rank-sum (Wilcoxon, 1992) test because the comparison includes different sets of cells. We used exclusively one-tailed tests chosen based on the difference of medians. The significance level was kept at 0.05 for figure annotations. There were 10 data categories often including data from the same cell but with a different recording mode (voltage clamp at either the excitatory or inhibitory reversal potential). We reinforced comparisons across experiments (algorithm versions) by pooling data and assessing the discriminating statistic over all. Thus we were interested in the comparisonwise error rate and did not need to adjust for multiple comparisons because we did not claim that one algorithm was better than another on the basis of an individual category (Bender and Lange, 2001). We highlighted individual comparisons on categories across and within algorithm versions to make specific claims about those categories. The notion of significance being open to interpretation (Bender and Lange, 2001), we elected to simply publish the P-value as well as the number of categories and let the reader make the final judgment. The effect size is the simple difference formula (Kerby, 2014), chosen because it is a normalized measure weighting the median difference between two matched sets against the size and frequency of cases that contradict the median difference and helps evaluate the meaningfulness of a judgment about significance.

We used two measures of performance, the fraction of trials for which the co-occurring stimulus was correctly predicted, the correct classification rate "CCR" and also the F_1 score

(Sokolova and Lapalme, 2009). The correct classification rate is the fraction of data points that were assigned the correct label. In our case labels were the ordinal rank, or index, of a drifting grating property as the stimulus label. The F_1 score is widely used throughout machine learning. It applies to binary classification and is useful in cases with skewed class sizes. This makes it a useful measure of discriminability for individual stimuli. The F_1 score is defined with respect to a specific stimulus label. $F_1 = 2 \cdot T_P / (2 \cdot T_P + F_P + F_N)$ where T_p are the number of true positives regarding that stimulus label, F_P are the false positives, and F_N are the false negatives. This value ranges between zero and one. We used the average F1 score across all stimulus labels as the objective function error measure because it penalizes cases where the classifier learns to predict one or two stimuli correctly at the expense of predicting others. The CCR is not resilient to this error but is familiar to a broader audience.

3.2.5 Characterizing Recorded Responses with Deflection and Definition of Epoch

All experimental signals were recorded at 20 kHz and downsampled to 1 kHz by means of 20 ms averaging. When removing spikes from membrane potential recordings the downsampled signals were further processed with a median filter with a five millisecond window. Simulated data was produced at 10 kHz and downsampled to 1 kHz by 10 ms averaging.

We calculated deflection by obtaining a baseline and subtracting that from the average signal during an epoch of choice (Figure 3.2a). The baseline was found for individual trials

by identifying a period of minimum variance that was 100 ms long and had a start time between 200 and about 100 ms before the onset of stimulus response. No baseline ever included a portion of a stimulus response in its estimation. The exact timing that the monitor displayed the stimulus was the "on" timing in Fig 2a. It was not recorded, instead we made a conservative estimate to mark a time before any stimulus response was likely to reach V1 given retina to V1 latencies (see below). For the purpose of stimulus discrimination we tested different lengths and positions for interval used to define deflection. We used a 166 ms window which contained the peak in average response during the epoch of choice (see below) and began either 66 ms before the peak, or halfway between the peak and the epoch start time if the peak was within 66 ms of the epoch start time. The deflection for each trial was the average value of the baseline subtracted from the average value within this window. This formulation gave the highest ability to classify stimulus by using the deflection. Therefore this was the fairest definition of deflection to use for comparison to dynamical discrimination. In previous work with this same data, the term "response" was used (Adesnik, 2017). The key difference being the length and location of the second window. The researchers integrated the difference from baseline without dividing by the length (i.e. it was not a mathematical average but proportional). This was tested and compared against our use of deflection. The absolute value of deflection was used in analysis.

We identified the timing of stimulus onset as follows. For each cell and type of recording (excitatory current, inhibitory current, and membrane potential) we averaged all recordings and applied a 50 ms running average to the mean recording. We then identified

the largest extrema in the first half of the mean recording. The mean recording was then binned by 50 ms intervals and a first order derivative estimate computed. The largest derivative immediately prior to the extrema denoted the bin in which the response began on average. Because it takes 70 -150 ms for activity to propagate from the retina to V1 L2/3 neurons (Gao et al., 2010) we subtracted 100 ms and rounded to the nearest half bin-width, 25 ms. This was finally defined as the "on" timing and occurs well before stimulus response. Therefore we captured the full response even allowing for variability in time of onset. The length of the stimulus presentation also varied, from 500 ms to 1000 ms. So the "off" timing also varied.

The response to stimuli is commonly categorized into three distinct epochs with much study and debate about their role in sensory processing (Phillips and Singer, 1974; Singer and Phillips, 1974; Duysens et al., 1996; Müller et al., 2001; Mazor and Laurent, 2005; Maravall et al., 2007; Gutnisky and Dragoi, 2008; Liang et al., 2008; Berens et al., 2012; Solomon and Kohn, 2014; Clawson et al., 2017; Bondanelli and Ostojic, 2018): the "on response" coinciding with the activation of a stimulus, the "steady-state" response which captures what follows the on response while the stimulus is still active, and the "off response" which is a widely observed perturbation or lingering effect after the stimulus has ceased. We did not know which epoch would allow the best discrimination. Some evidence pointed toward fast attractor dynamics in the on response (Berens et al., 2012; Miconi et al., 2016), but we treated stimulus response epoch as a hyperparameter and captured differences in the performance of our dynamical discrimination algorithm (see appendix A.2). Referencing both the on and off timings as t_0 and t_f respectively we defined the "on

response" as [t_0 ,t_0+250 ms], the "full response" as [t_0 ,t_f +250 ms], and the "off response" as [t_f - 70 ms,t_f + 250 ms]. The discriminability of trajectories from these three epochs have some scientific merit in their own right and are reported in the appendix A.2. These choices allow us to examine performance under different circumstances and thereby optimize classification performance by using domain specific insights.

3.2.6 Single Neuron Modeling

We used models of a single-compartment Hodgkin-Huxely type neuron and a morphologically complex pyramidal neuron implemented with the NEURON environment. The morphologically complex pyramidal neuron was developed for other research (Palmer and Stuart, 2009) and extensively explained. It is available for general use from model dB (Anon, n.d.). It was modified to allow experimental manipulations of the number of synaptic spines distributed across the various compartments and having one synapses for each spine and to allow each synapse to be driven with independent signals. Consult our model data generation scripts for model form and parameter details.

For both neuron types we rescaled and resampled the X dimension of the Lorenz system. The Lorenz system (Lorenz, 1963):

$$\begin{aligned} \frac{dX}{dt} &= \sigma(Y - X) \\ \frac{dY}{dt} &= X\,(\rho - Z) - Y \\ \frac{dZ}{dt} &= XY - \beta Z \end{aligned} \qquad (3.1)$$

was integrated using MATLAB's ode45 for 10K time steps with a nominal step size of $dt=10^{-4}$, initial conditions were randomly chosen and uniformly distributed between [-16,16] for the X and Y dimensions, and [-56,56] for Z. After generating data the X dimension was kept and the other dimensions were not. The Hodgkin Huxley equations act as a low-pass filter and our single-compartment model distorted input current oscillations above 333 Hz. Therefore if the integral of the squared absolute value of a fast-fourier transform of the signal from 0 to 300 Hz accounted for less than 90% of the total integral then the trial was resampled $X'(t) = X(t/\tau)$. Resampling was done by regenerating new Lorenz data with a false time step $dt'=dt/\tau$, where τ (in units of 10^{-4} s) was chosen such that 90% of the signal power was in Fourier modes below 300 Hz. The process was repeated until a set of initial conditions and τ were found that satisfy our acceptance criteria. Let $\mathcal{X}$ be the Fourier transform of X in the frequency domain, then:

$$\tau = \frac{(0.9\int_0^{\infty}|\mathcal{X}(f)|^2\,df)}{(\int_0^{300}|\mathcal{X}(f)|^2 df)}. \qquad (3.2)$$

We treat the ρ parameter as a latent variable we were attempting to identify and vary it between the integers [20,40]. This lets us test whether we can predict fine changes in a dynamical parameter (ρ) even after transforming the dynamics into the membrane potential of two classes of model neuron. Our method of attractor reconstruction, delay embedding, is robust to arbitrary projections of a dynamical system onto one dimension. A weighted adjacency matrix typical of network model is an example of such a projection, except for the additional transformation from a continuous time series to a discrete point

process which is often used. Nonetheless it has been shown that a spiking network can encode Lorenz attractor dynamics which can then be viewed by projecting high-dimensional population spiking onto a lower dimensional state-space (Eliasmith, 2005). Therefore it was sufficient to project the Lorenz dynamics onto a single dimension and stochastically encode that dimension with discrete events if we want to study the representation of network attractor dynamics by single neuron inputs. We did this while taking the ρ parameter through values on either side of hopf bifurcation at $\rho = \frac{\sigma(\sigma+\beta+3)}{(\sigma-\beta-1)} \approx$ 24.7and into regimes where initial conditions often result in chaotic dynamics.

For the single-compartment neuron we rescaled $X'(t)$ to be within a realistic range for current units [-0.15 nA, 0.15 nA], $I\ (t) = \alpha_1(X'\ (t) - \zeta_1)$. The scaling factors ζ_1 and α_1 were fine tuned by a loop which adjusts them to produce a desired mean spike rate for a group of 210 trials (21 values of ρ each with 10 initial conditions unique to each value of ρ). This produced a range of spiking values and was repeated until there were at least three trials for each ρ value for each desired number of spikes [0,20] per 1 second trial. This gave us multiple trials for each value of ρ at each level of spiking we were interested in testing. This method for generating the desired amount of spiking was inexact but preferable to search algorithms that precisely controlled the spiking in every trial. These algorithms took a long time to find solutions and often found undesirable solutions, such as scaling the inputted current to have a standard deviation near zero. Consult our model initialization scripts for model details.

For the complex morphological neuron we rescaled $X'(t)$ to be consistent with an instantaneous event probability for an inhomogeneous Poisson process: $P(t) = \alpha_2(X'(t) - \zeta_2)$. The factors ζ_2 and α_2 were further fine tuned the minimum and maximum value to get a desired mean spike rate for a group of 210 trials, as with the single-compartment model. Initially we set the range and maximum value to be 3×10^{-3}. We vary the number of synapses between 80 and 300. For each synapse the probability of firing at any time was $P(t)$. Thus larger $P(t)$ values resulted in greater synchronicity among synaptic events. Synaptic transmission was modeled with NEURON's Exp2Syn function which is a two-state synapse with a rise time of 0.2 ms and a fall time of 1 ms. The peak synaptic conductance, g_{max} (units of μS) was a function of the number of synapses N_{syn} such that if K% of synapses were active the total peak conductance was independent of N_{syn}: $g_{max}(N_{syn}) = g_0(80/N_{syn})$ where $g_0 = 5\times10^{-4}$ μS.

We employ modeling to test the effect that realistic sources of error may have on the ability to infer the dynamics which underlie neural inputs. The two key variables that we controlled for were spiking and synapse numerosity in an extended multi-compartment model. To that end we defined 15 logarithmically increasing bins to contain model trials with similar levels of spiking. These bins started with [0,1), ended with the 14th bin [39,50) and the 15th bin [50,∞). Since the single compartment neuron did not feature synapses these bins specified all control categories. Not all bins were populated. Each bin contained multiple trials for each value of ρ . If there were fewer than three trials for a value of ρ then those trials were ignored and not analyzed. If there were fewer than 45 total trials in a spiking rate bin after removal of underrepresented ρ values, that bin was ignored and none

of the trials in that bin were analyzed. As a result there were 12 spiking categories that were analyzed for single-compartment neuron data. The three unfilled bins were for trials with more than 20 spikes, which was our minimum standard when generating data. The morphologically complex multi-compartment model neuron has two variables to control for: spiking and synapse numerosity. For each spiking bin we defined a number of synapse bins to categorize trials in two dimensions (spiking and synapse numerosity). There were more model trials with low numbers of spiking due to the inexact way we generated the desired amount of spiking (by the average spiking rate of 210 trials). Because trials with a small amount of spiking were overrepresented it was advantageous to define more synapse bins for the spike bins containing fewer than five spikes per trial. Thus there were 19 synapse bins for each of the first five spike bins going from 80 synapses to less than 125 in steps of 10, then 125 to 305 in steps of 15. There were 16 synapse bins for the remaining spike bins, going from 80 to 305 in steps of 15. Thus there were 255 possible bins and 83 were accepted for analysis. Again each bin had to contain more than 45 total trials and each ρ value had to have more than two trials to be included. In the main text we plot a sampling of these categories, attempting to show how spiking impacts performance while keeping synapse numerosity approximately constant (Fig 5e). All the resampled Lorenz dynamics inputted to each cell were preserved without rescaling and analyzed together as one group.

3.2.7 Dimensionality Expansion

The version of time-delay embedding we perform is described in (Brunton et al., 2016, 2017), a short summary follows. First we concatenated the epochs of interest from all recordings from the same cell to make a time series with length T. Next, we chose a delay

time below or at the approximate smallest relevant time-scale, *dt*, (1 ms in our case) and a number of times to repeat the delay N_d=100 such that the N_d^{th} delay was longer than or at the largest relevant timescale (100 ms) . We then created a data matrix with N_d+1 rows, each of which were time shifted copies of the data with length T-dt·N_d (called a Hankel matrix). Then we performed singular value decomposition on this Hankel matrix (we tested other dimensionality reduction algorithms). The principal components were now the dimensionality expanded version of the data. Thus we simply over-expanded with time-delays then used dimensionality reduction to go back down to moderate dimension. There was no clear cut off in the eigenspectra so we tested keeping between three and seven components by running each choice through the analysis program. We did not Z-score the rows of the Hankel matrix but did shift the trajectories resulting from dimensionality reduction such that the mean point of the entire set of trajectories was at the origin. A delay embedding is guaranteed by the Whitney-Takens delay embedding theorem to be able to reconstruct a D dimensional state state space from a one-dimensional recording by taking no more than 2D+1 delays of the recording and plotting them against each other (Kostelich and Schreiber, 1993). We call this process "dimensionality expansion" to provide the intuition that if an analyst can do something after dimensionality reduction of high-dimensional data they can at least attempt it on one-dimensional data too. The analyst would simply use SVD on any over-embedding such as a spectrogram or such as a Hankel matrix like we have demonstrated.

3.2.8 Maximum Likelihood Estimation

It's possible that dimensionality expansion alone can improve discriminability between stimulus responses without appeal to governing dynamics. This would work because different population responses were projected onto somatic responses in a way that overlaps. Hence deflection in the original time series confuses these factors but dimensionality expansion may re-separate them. We tested measures of deflection in each dimension of the expanded trajectories but this did not significantly improve classification. An alternative is to examine whether different trajectories prefer to spend time in different regions of phase space. We used the performance of a Maximum Likelihood Estimation based classifier to quantify this separation (see appendix A.4).

Since trajectories in three dimensions tended to form oscillations around a long axis, cylindrical coordinates (axial z, radial r, and angular θ) were a natural way to describe them. We partitioned axial and radial coordinates into discrete bins and ignored the angular coordinate then counted the number of time points that coincided with each bin. Thus each trajectory was described by a histogram of its axial cross-section. We collected the cross-section histograms associated with each stimulus, using 75% of the examples for each stimulus. Then we employed a two dimensional kernel smoothing density to get a two dimensional probability density map for each trajectory $M_t(z,r;i)$ where i is the trajectory index (Figure A.9a) and for each stimulus $M_s(z,r;j)$ where j is the stimulus index (Figure A.9b). Each point (z, r) identifies an axial-radial bin and was assigned the probability that a trajectory time point selecting at random will be in its bin. This gives a set of probability maps that show the probability for a trajectory to occupy a region of phase space (a cross-section bin) dependent on each stimulus. Thus to test whether one of the 25% of

trajectories we held out coincided with the presentation of a particular stimulus we used the probability maps to calculate the joint probability of observing all the time points given that stimulus (treating each point as independent), $P\ (i;j) = \prod_z\ \prod_r\ M_s(z,r;j) \cdot M_t(z,r;i)$, for practical application we used the log likelihood $L\ (i;j) = \sum_z\ \sum_r\ -(log\ (M_s(z,r;j)) + log\ (M_t(z,r;i)))$ The stimulus whose probability map yields the highest joint probability was the stimulus with the highest likelihood of co-occurring with the trajectory and thus was the classification, $C_j = \underset{i}{\text{argmax}}\, L\ (i;j)$. By repeating this process 510 times with a different hold-out set each time we can gather sufficient statistics to gauge whether this prediction method was effective.

A detail essential for reproducibility is that the trajectories co-occurring with each stimulus show some displacement between their central axes and central points. Thus in order to get joint probabilities one must subtract the central point of each stimulus-associated trajectory set before computing $M_s\ (z,r;j)$.

3.2.9 Reliability Tuning Curves

A tuning curve is defined as mean deflection in response to stimulus. Therefore we recorded the average evoked deflection for each stimulus, D_i, where i is the stimulus index out of N stimuli. For a single cell and intracellular recording method, the stimulus that evokes the largest deflections on average is the most preferred stimulus, and the one that evokes the smallest deflection is the least preferred stimulus. The least preferred stimulus is not necessarily the stimulus that is least similar to the preferred stimulus (i.e. not the

anti preferred stimuli). Since stimuli were continuously and monotonically varied along one parameter (orientation, contrast, size) we used the ordinal number difference between two stimuli as a similarity metric. Stimuli were indexed by an ordered ranking. Thus stimuli numbers 1 and 6 would be the smallest and largest drifting grating if size were varied and they would also be maximally dissimilar, while 5 and 6 would be the largest and second largest and maximally similar. We defined reliability as the mean deflection in response to the same stimulus divided by the standard deviation of deflections in response to that same stimulus, R_i where i is the stimulus index out of N stimuli. In order to measure the correlation between reliability and similarity to either the most or least preferred stimulus we defined a least/most similarity function, S_i as follows. The ordinal values of the most and least preferred stimuli were recorded. For all stimuli, the absolute value of the ordinal difference between themselves and the least preferred stimulus was noted, then the same was noted for the most preferred stimulus. The smallest of these two absolute ordinal differences was kept. Thus for each stimulus we have recorded the absolute ordinal difference between itself and either the most or least preferred stimulus (depending on which difference was smaller). We then divided by the number of distinct stimuli (either six for size or contrast, or eight for orientation). The least/most similarity function was then: $S_i = min\left(\left\{\left|i - \underset{0<j\leq N}{\operatorname{argmin}} D_j\right|, \left|i - \underset{0<j\leq N}{\operatorname{argmax}} D_j\right|\right\}\right)/N$. We also have the reliability measure as defined above. We measured the Pearson Correlation between reliability, R_i and this least/most similarity score S_i. Because this measure was applied individually to each cell it quantifies whether the U-shaped trend in Figure 3.4b was a property of cells individually. If D_i tended to be proportional to R_i for some cells and inversely proportional

for other cells then Figure 3.4b may still appear U-shaped but there would be no correlation between R_i and S_i.

We also explored possible causes for changes in reliability. These include the coefficient of variation of a response and the mean of normalized residuals. The mean of normalized residual is a measure of noise. To calculate it, we first calculated the average response of a stimulus to all repetitions of a stimulus. We subtracted this mean response from a single trial, the result was a residual time series. To normalize the residual time series, we divided by the mean response time series from the first step. Therefore each point in the new time series was the signed fractional error between the single trial and the mean response. The average value of this time series was defined as the mean normalized residual. We also examined a variant where we computed the absolute value of the normalized residual time series before computing the average. For any given recording we calculated its normalized residual by comparing to the other responses to the same stimuli. We compared noise with fluctuation size. We defined fluctuation size as the coefficient of variation, which is the standard deviation of a timeseries divided by the mean value of the same time series. Thus both measures control for average deflection. These measures let us test whether decreased noise or decreased fluctuation size was responsible for increased reliability.

3.2.10 Genetic Modification of SINDy

We used the SINDy algorithm because it is well supported and amenable to modification (Brunton et al., 2016). In short, we pre-computed the derivatives, dV, of a singular value decomposition base time-delay embedding (Brunton et al., 2016, 2017) (dimensionality

expansion),V, of a single time-series. We then pre-computed many polynomial combinations of the original data (including a constant term) Θ. A dynamical system was therefore captured by a matrix Ξ projecting the polynomial combinations onto the derivatives $dV=\Xi^T\Theta$ .The critical insight is to set most of the elements of Ξ to zero so that the dynamical system is readable, tractable, and generalizes to the rest of state-space. Originally (Brunton et al., 2016), elements of Ξ were chosen to be non-zero by identifying thresholds through hand tweaking such that only Ξ elements exceeding the thresholds were included in the fit. Instead we used a genetic algorithm to automatically decide which elements to set to zero without a threshold. For simplicity we call this "genetic SINDy". Note that if one specifies the locations but not the values of nonzero elements with a binary-valued bitmask Ξ matrix, ${}^B\Xi$, then a ${}^B\Xi$ is a template which can generate diverse kinds of Ξ matrices because fitting the coefficients specified by ${}^B\Xi$ to two different trajectories would produce two different Ξ. The original work showed that using time-delay reconstructions of undersampled systems yields Ξ matrices that are characteristically non-sparse in the last dimension (Brunton et al., 2016). The original work also noted that it requires long periods of time in diverse situations to capture the best invariant models of the system. This is a characteristic we exploited to get local approximations instead of the invariant models the original work sought to obtain.

In order to avoid numerical error one must normalize Θ. A key difference is that the original paper (Brunton et al., 2016) divided by the L^1 norm but we had better results by Z-scoring (Hastie et al., 2009a) each variable (row of Θ) and always including a constant term in Ξ. This forced information about the average variable value into the constant term, making it available for the classifier in later stages. We tested the inclusion of second order

derivatives, d^2V/dt^2, and inclusion of more dimensions, as well as numerous other variations but settled on three dimensions and first order derivatives, dV/dt as giving the highest utility with the least complexity and compute time (see appendix A.2). Another departure from the original implementation of SINDy (Brunton et al., 2016) was the addition of a single three time-step smoothing window after estimating derivatives using the fourth order method. We did not rigorously compare the inclusion of non-polynomial forms as performance was good enough that we could test our hypothesis without the additional complexity.

A genetic algorithm must be initialized with a very diverse population of individual "guesses" at a solution to the problem. In our case an individual solution was a bitmask matrix, ${}^{B}\Xi$, the same size as Ξ, but consisting only of zeros and ones. Ones marked the location of Ξ elements to keep as non-zero when creating an ODE model in later steps. To get an initial population we used an unsupervised threshold method to decide which elements of ${}^{B}\Xi$ were one and which were zero (included with our software). We treated each trajectory individually and used bisection search to find the largest threshold (for each column of ${}^{B}\Xi$) that resulted in at least one non-zero element. This gave a maximally sparse representation and largely reproduced the findings reported elsewhere (Brunton et al., 2016) when used on a fully sampled Lorenz system. The result was a set of unique ${}^{B}\Xi$ that was no larger than the number of trajectories.

We used a "mating" (crossover) process to create 300 unique individuals. For each ${}^{B}\Xi$ in this set we obtained Ξ matrices for each trajectory. Twenty five percent of the trajectories

were held out for testing, the remainder were used to find the coefficients of the system of equations. After testing fitness the best 45 $^{B}\Xi$ were "mated" and "mutated" to generate 300 new forms of the equation. A "sexual genetic algorithm" requires a method for combining possible solutions, a "mating" process. This involves three steps, selecting individuals to combine, deciding which attributes to keep in the "offspring", and a way to mutate the offspring. We kept the best 45 unique $^{B}\Xi$ and ranked their performance (worst is 1 best is 45). We then used 255 tournaments to select parents, mate them, and produce 255 new $^{B}\Xi$ matrices. One parent was selected by cyclically stepping by one through the best 45, the second parent was selected at random with a probability in proportion to its rank. A parent was not allowed to mate with itself. The nonzero elements of children were selected by keeping elements which were nonzero in both parents and with probability one-half if it was nonzero in only one parent. Finally, mutation was implemented by flipping one or more randomly selected elements to its opposite value. Each element of the child matrix was subject to mutation with a specific probability called the "mutation rate". The mutation rate was 0.15 to produce the initial population. It was set to 0.05 for the first generation and was periodically halved until it was set to zero for the last ten percent of generations. The number of halvings depends on the initial mutation rate, r_{mut},and the number of elements in the Ξ matrix, N_{Ξ}, according to $N_{mut} = \lceil log_2 (r_{mut} \cdot N_{\Xi}) + 1 \rceil$and was not allowed to be smaller than 2. We tested other methods for mating on small fractions of original and simulated data, including: transferring columns or rows to the children intact, selecting half of the elements from each parent (either at random or in a structured way), or simply keeping all elements which occur in any parent. The choice of mutation rate and

halving periods, as well as mate selection, mutation methods, the population size, and fraction to keep were selected by hand tweaking on fractions of data and simulated data.

The number of generations to run the algorithm increased by 100 for every three columns of the Ξ matrix (which may include second derivatives as well as higher dimensions). If the errors of the 45 ${}^{B}\Xi$ matrices were identical or within one one-thousandth of the range of errors in the initial population then the algorithm was terminated early. This never happened when the objective function for the algorithm was classification ability and happened only rarely when the objective function was goodness of fit. The number of generations was tested by hand to be long enough to ensure convergence but not long enough to produce over fitting.

In our case different ${}^{B}\Xi$ represented different possible solutions and we had three objectives to consider. We desired a Ξ matrix which can be fed into a classifier and perform well, we desired that this Ξ matrix be sparse (to avoid overfitting and improve interpretability) and last we desired that the Ξ matrix describes a good model of the dynamics. The objective function for classification ability was noisy because the hold-out set was small for the generation updates, (see below), therefore we retested the best 45 at each generation.

3.2.11 Classifier Objective Function

An objective function accepts a ${}^{B}\Xi$ matrix and outputs a scalar value which is lower for ${}^{B}\Xi$ that are better at satisfying some objective. For predicting the stimulus based on the

coefficients of a fitted ODE the genetic algorithm objective function started by fitting Ξ matrices to each trajectory using only the coefficients specified in the ${}^{B}\Xi$ matrix it accepted as input. Next, For each stimulus, 25% of fitted Ξ matrices were held out for cross validation. A random forest (Breiman, 2001; Sokolova and Lapalme, 2009) was then trained on the remainder (the 75%) and tested on the cross validation set (the 25%). This was repeated 10 times, selecting a different twenty five percent each time (this is sometimes called hold-k cross validation with bootstrapping (Arlot and Celisse, 2010)). The classification performance was the average F_1-score for all stimulus labels (Sokolova and Lapalme, 2009). We subtracted this value from one such that good performance was a low number that still ranges between zero and one, and constituted 80% of the objective function value. The other 20% of the fitness value was a regularization term: the fraction of possible terms which were nonzero (i.e. sparseness).

3.2.12 Goodness of Fit Objective Function

For finding a set of coefficients (specified with ${}^{B}\Xi$) that allowed the highest quality ODE model, the objective function started by fitting Ξ matrices to each trajectory using only the coefficients specified in the ${}^{B}\Xi$ matrix it accepted as input. Next it used all the original points on the trajectory as initial conditions to integrate the fitted ODE four timesteps. Goodness of fit was $1\text{-}R^2$ (the coefficient of determination) between the derivatives predicted by integrating the ODE and the derivatives of the data shifted by five timesteps. A sparseness regularizer was used such that goodness of fit was 80% of the objective function value and sparseness was 20%.

3.2.13 Ensemble Classification and Out-of-Sample Generalization

The genetic algorithm tested one ${}^{B}\Xi$ matrix at a time, however each matrix was pulling out a different set of coefficients and therefore might have been highlighting different dynamical attributes. It is often found that an ensemble of independently trained classifiers can cooperatively vote on a classification and that doing so often cancels out bias that cropped up during the training of any individual (Hastie et al., 2009b). We found that ensemble methods decreased overfitting tendencies for dynamic discrimination. Therefore, the 45 ${}^{B}\Xi$ unique matrices which were best in the final generation of genetic SINDy voted on making a final classification to complete our process called "dynamical discrimination". When we were training random forest (Breiman, 2001) regression trees instead of classifiers we used the median of the ensemble. For classifiers we used the mode with ties broken by choice with the highest ranked ${}^{B}\Xi$ according to final generation fitness.

Because the cross validation process was repeated once every generation there was information leakage and over-fitting effects were possible for the algorithm as described. To measure overfitting we re-ran the entire genetic algorithm and ensemble process on data where the stimulus labels were scrambled. Performance on random surrogates was stable, despite high variability in the performance on the original data (Figure 3.3b). Random surrogate performance was indistinguishable from chance for deflection and classification based on Ξ that were fitted without regard to classification performance (Figure 3.3a, 3c).

In order to report values without confusing overfitting for reproducible (out-of-sample) performance we encapsulated the entire genetic algorithm and ensemble process in another layer of hold out testing. We held out one example of presentations for each of the 6-8 stimuli and ran the entire algorithm, then predicted the stimuli for the examples which were never used for training. This was repeated twenty times and the final values reported in the main text and in Figure 3.3d, are the results of this 20-fold hold-one-out test. Using the data from the cell with the most trials, we verified that the final holdout did not perform better than chance when trained on random surrogates. Therefore the final holdout performance is a valid out-of-sample generalization.

3.2.14 Hyperparameter Optimization

Augmenting Sparse Identification of Nonlinear Dynamics (SINDy) with a genetic algorithm to provide a representation of time-series suitable for a random forest classifier requires many choices which can affect outcomes. Research that makes scientific comparisons but does include an account of hyperparameter optimization cannot be adequately reproduced, or checked for bias. Hyperparameters are the choices an analysist makes about a machine learning algorithm that effect the performance of the algorithm, but which the algorithm itself cannot tweak or adjust. For example, a researcher may train a classification algorithm to predict one variable called the “class label”. The researcher must provide other variables called the “predictive features” which are supposed to serve as the basis of comparison. The researcher might access to a great many variables that might serve as predictive features and be forced to choose just a few, Thus the choice of *which* variables to include is a hyperparameter. Alternatively, the algorithm may be equipped the ability to choose

which variables to include on its own, but must be told how many variables to include. In the latter case the *number* of predictive features is a hyperparameter. Every algorithm has a set of hyperparameters, sometimes the choices are obvious but when they are not they must be tested in a process called "hyperparameter optimization". With respect to SINDy performed on delay-embedded (dimensionally expanded) data one must decide what size of delays, how many delays to include prior to dimensionality reduction, and what dimensionality reduction algorithm to include as well as a slew of ODE related choices such as what polynomial order to include, whether to include other non-polynomial forms such as sinusoids or sigmoids, whether to include a quenched-noise driving term (treating extra dimensions as a time-varying input) (Brunton et al., 2016), how many dimensions to include, how to get stable numerical estimates of derivatives, whether to use higher-order derivatives, how to normalize the data, and preprocessing steps such as detrending, and filtering and what period of the timeseries to fit. A genetic algorithm requires even more choices such as methods for mate selection, crossover, mutation, as well as what terms to include in the objective function, how to weight those terms, and stopping conditions. Even the selection of a classifier algorithm to perform the last step presents a set of choices that can alter a scientific comparison.

Few of these choices can be made a-priori, and there were too many to test completely. Choices such as these are known as "hyperparameters". They must be reported on to demonstrate that all effort was made to maximize the performance of machine learning algorithms before making claims about comparisons. Most choices were made by hand testing on small fractions of data or on simulated data. Some choices made little difference,

some made the algorithm worse, others could not be justified due to the exorbitant computational time required. These choices are important to be aware of for replicating results but are not extensively reported on here, except to state them: we chose 100, one millisecond delays and chose Singular Value Decomposition (SVD) to reduce from 100 to between three and seven dimensions (and extensively tested this narrowed range). We included only polynomial terms up to the third order (because it is the same order as a FitzHugh-Nagumo ODE), and classification did not benefit from including an additional dimension as a quenched-noise driving term. Derivatives were estimated by using a custom algorithm based on standard fourth order derivative methods (Brunton et al., 2016) but added smoothing (window size three timesteps) as a last step. Derivative estimation methods were chosen to ensure the derivatives matched the trajectory when accumulated. Data were not detrended but were downsampled to 1 kHz. After dimensionality expansion the entire set of embedded trajectories was centered at the origin. The choice of time period, number of dimensions and derivative order were tested with a more exhaustive optimization approach. For the genetic algorithm, mate selection, crossover method, and mutation method, as well as the size of the population and the fraction to keep at each generation were all tested on small fractions of the data (and choices stated above). The choice of regularization factors and terms to include in the objective function were narrowed by hand on a small fraction of data and then a few remaining options were exhaustively tested. There were 51 trees in the random forest algorithm for final results and 5 trees for evaluating the classifier objective function in intermediate generations of the genetic algorithm. Alternative methods of classification, such as fitting ODEs to the trajectories co-occurring with a specific stimulus and then assessing which ODE best-fit a

test trajectory, yielded such poor performance or additional complexity that they did not justify inclusion in this paper. The inclusion of additional constraints such as a goodness-of-fit constraint in a classifier objective function did not perform well enough to justify the testing needed to find the optimum choice. It was found that an early termination stopping condition based on convergence to a narrow range of error values achieved the same results as a stopping condition based on a lack of diversity among possible solutions. The choice of 100 generations for every three columns of $^{B}\Xi$ was made by hand testing.

Some choices were selected for exhaustive testing because they either had scientific value: period of stimulus presentation to train on, inclusion of second order derivatives, and how many dimensions to include. Others were chosen for exhaustive testing because the effect was complexly related to other factors being tested: regularization factor, choice of dimensionality reduction (SVD versus independent component analysis), and whether to use an ensemble method. These were tested by running genetic SINDy on either all of the data or just on I_{LO}, or I_{LS} and comparing the cross-validation of the final generation with the same for random surrogate data. The parameter set with the best classification ability with the least variability and least overfitting was selected. This was three dimensions with a regularization weighting factor of 0.2 and only first order derivatives, using an ensemble classification method with SVD for dimensionality reduction and the period coinciding with the onset of stimulus response was the most informative. These testing results are reported in appendix A.2.

3.2.15 Integration of Ξ Matrices

Because we normalized Θ components by Z-scoring (Hastie et al., 2009a) them we had to carry out a change of coordinates at each step when integrating the ODE models to assess how they captured trajectory details. Let $\mathcal{E}$ be the function that creates polynomial combinations of a trajectory, V, such that $\Theta=\mathcal{E}(V)$. When integrating the ODE to create a simulated trajectory, $V'(t)=dV'(t-1)+V'(t-1)$, the derivative term at each timestep becomes $dV'(t)=\Xi^T[\mathcal{E}(V'(t-1))-\mu_\Theta]/\sigma_\Theta$, where μ_Θ and σ_Θ are the mean and standard deviations of each column of the Θ used to estimate Ξ from the experimental data.

We used points from the initial trajectory and found that our models tend to be difficult to integrate. Even with a stiff ODE solver initial conditions frequently "blew up" wherein the derivatives became very large, or the derivatives rapidly extinguished. Therefore we tested multiple initial conditions and plotted the ones that produce trajectories which remain in the neighborhood of the original trajectory for as long as the original trajectory was, and that explore a volume similar to the original trajectory. This was done by rejecting initial conditions that produce trajectories whose standard deviations (along each dimension) were all less than five times the same standard deviations of the original trajectory and exceeded one twentieth of the original trajectory.

3.2.16 Linear Stability Analysis

For non-linear dynamical systems such as those approximated with SINDy the behavior in the vicinity of a fixed point is often analyzed through the eigenvalues of a Jacobian matrix evaluated at the fixed point. A Jacobian matrix is the matrix of all partial derivatives with respect to the main variables. We can evaluate this using the chain rule and our

normalization factors. Normalization factors included a translation as well as a rescaling. Using Matlab's symbolic toolbox we solved the ODEs for coordinates where the derivatives became zero. The translation did not affect the Jacobian but did need to be accounted for when solving for fixed point coordinates the same way it was accounted for when integrating (see above). We report on all real-valued fixed-points. We evaluated the Jacobian for the best one of the 45 Ξ matrices according to their fitness values on the final generation of the genetic algorithm. We evaluated all fixed points for each trajectory individually. If the eigenvalues of a fixed point have an imaginary component the dynamics are locally oscillatory, if the real component is positive they diverge away from the fixed point, if the real component is negative they converge towards it, if the real component is zero the dynamics form a cycle. If the number of fixed points changes when a system parameter (such as stimulus label) is changed then that parameter is said to take the system through a bifurcation, likewise if the sign of the real component of the maximum eigenvalue of the Jacobian changes. We report on the number of fixed points and the eigenvalues as a function of what stimulus was presented in appendix A.3.

3.2.17 Data and Software Availability

The data, data analysis software, and modeling software used in this study are available from the author, James K Johnson, upon request or by searching for the project titled: "Stimulus tuned dynamical trajectories underlie synaptically driven transmembrane signals in visual cortex" on the Open Science Framework OSF.io once the companion manuscript of the same title has been accepted for publication.

3.3 Results

3.3.1 Dimensionality Expansion Captures Dynamically Rich Neural Trajectories from Single Neurons

Synaptically driven transmembrane electrotonic fluctuations contain rich information about network activity (Figure 3.1c) but it is not clear how to get that information. If behavioral responses to stimuli are consistent then fluctuations of neural activity following stimulus presentation should also be consistent at some level of abstraction. The concept of attracting sets from non-linear dynamical systems theory is the abstraction proposed by the attractor network computational paradigm (Amit and Amit, 1992; Eliasmith, 2005; Wu and Amari, 2005; Chambers and Rumpel, 2017). It describes brain activity as trajectories in a high-dimensional state-space and maps trajectory characteristics like shape or location to specific brain functions (e.g. memory, movement, or recognition (Daelli and Treves, 2010)). Near a fixed point most dynamical systems fall into a simple, or at least quasiperiodic, dynamical pattern (e.g. limit cycle oscillations, monotonic convergence) (Strogatz, 2018), and similar initial conditions produce similar trajectories over short timescales unless a dynamical bifurcation has occurred. Thus, if this perspective is valid, brief snippets from trajectories co-occurring with different brain functions will be different.

Classification ability is not epistemologically valuable but using it as a basis for comparison can be. Therefore recordings were placed into 10 categories according to signal type and stimulus characteristic (Figure 3.1c). Types of recording and stimulus characteristic are denoted with left-superscripts above I for voltage-clamp recordings and V for current-

clamp recordings (Figure 3.2a). The two categories $^{I,O}I$, and $^{E,O}I$ feature transmembrane current recordings of synaptic inhibition (^{I}) and excitation (^{E}) respectively, and drifting grating orientation (^{O}) was varied. The two categories $^{I,S}I$, and $^{E,S}I$ are the same but size (^{S}) was varied. For $^{I,C}I$, and $^{E,C}I$, contrast (^{C}) was varied. For the categories $^{R,S}V$, and $^{K,S}V$ the recording apparatus was in current clamp mode and size was varied and spikes were removed (^{R}) or kept (^{K}), respectively. For two more membrane potential categories contrast was varied $^{R,C}V$ and $^{K,C}V$. Figures show only data with spikes removed because of minimal difference in outcome. There are 20-121 recordings for each cell (median is 68) with 3-21 examples of each stimulus (median is 11).

To employ attractor networks in intracellular recordings from mouse primary visual cortex responding to visual stimulation, we used time-delay dimensionality expansion (Lainscsek and Sejnowski, 2015; Oprisan et al., 2015) of 500 ms recording snippets to project transmembrane current and potential fluctuations onto intermediate dimension neuronal trajectories (Figure 2b). The observed trajectories form oscillations confined to conical, cylindrical, or spherical regions (Figure 3.2c, A.9). Trajectories appear limited to regions of phase space according to recording context (Figure 3.2c, A.9). Patterns include nesting conic trajectories inside one-another, and displacement of oscillation centers. As evinced by Maximum Likelihood Estimation (MLE) (see methods, appendix A.4), trajectories were distinctive but not classifiable. Therefore we eliminated the simpler hypothesis that dimensionality expansion without dynamical systems characterization is sufficient for stimulus classification.

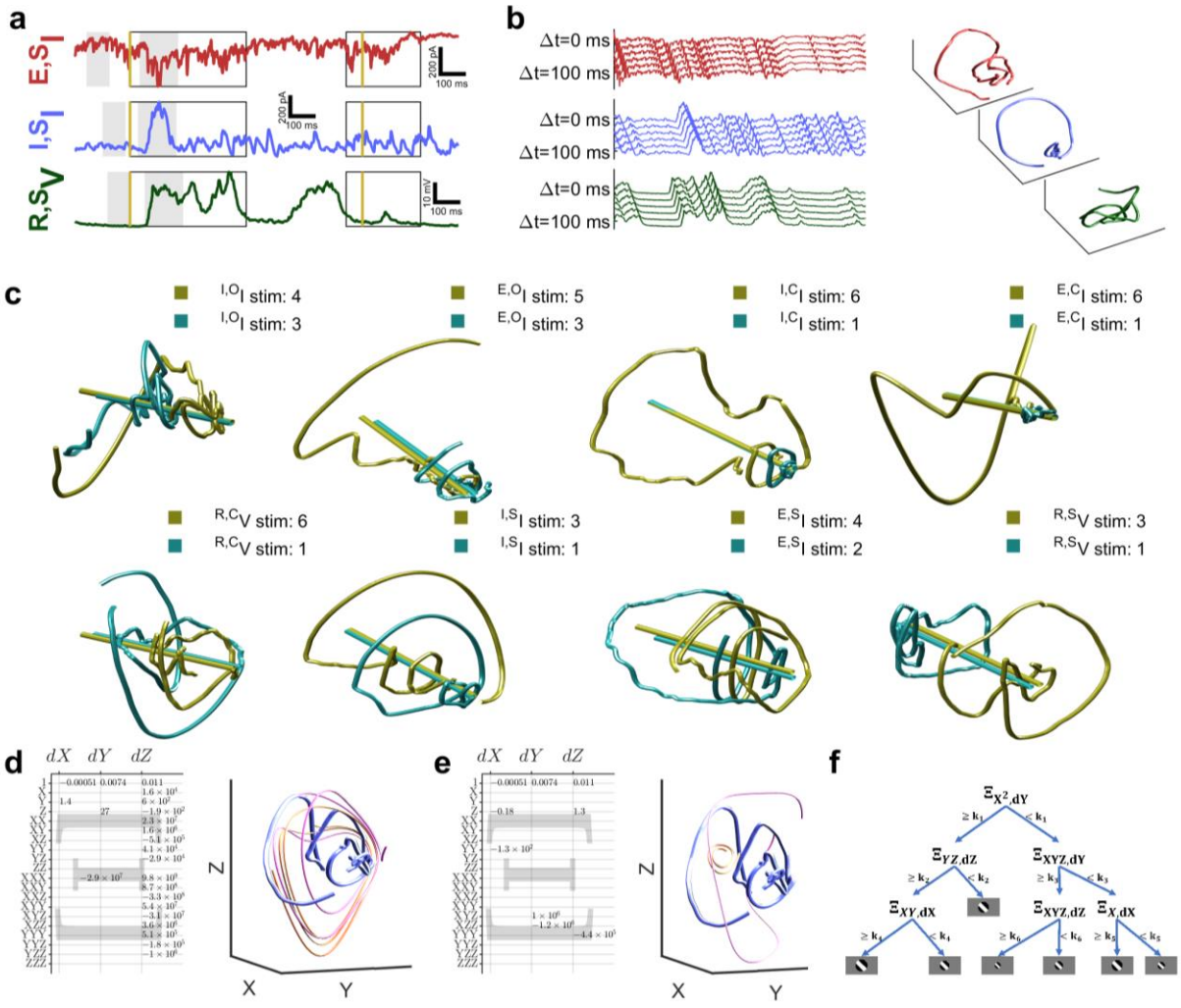

Figure 3.2 | Time-delay embedding of intracellular recordings reveals varied dynamical trajectories and equations are fitted to them. **a**, Excitatory (red) and inhibitory (blue) transmembrane current, and potential (green). Gold bars: stimulus on/off times. Outlined boxes: periods defining on response (early box) and off response (late box). Stimulus on time through the end of the off response defines the full response. Gray: periods defining deflection as the difference between early and late period means. **b**, Left column: Time-delay embeddings illustrated with every 20th 1 ms delay of recordings from panel a. Right column: Neural trajectories visualized after singular value decomposition of 100 1 ms delays. **c**, Two trajectories coinciding with the most (gold) and least (aqua) preferred stimuli (largest/smallest mean deflection respectively) for one cell from eight recording categories. Figure A.9b succinctly characterizes all trajectories. Central axes (gold/aqua bars) are parallel. Trajectory characteristics include: axial displacement of densest regions (see $^{R,S}V$, $^{I,S}I$, $^{I,O}I$), opposed directions of divergence ($^{R,S}V$), and nesting ($^{I,O}I$, $^{E,O}I$, $^{I,C}I$, $^{E,C}I$, $^{I,S}I$). Most cell's trajectories occupy conic regions (see Figure A.9), but some ($^{R,C}V$, $^{E,S}I$, $^{R,S}V$) are cylindrical or spheroid when plotted together. **d**, A Ξ matrix optimal for trajectory modeling. Left: Ξ coefficients. Right: A neural trajectory (blue) and reconstructions (magenta/yellow). **e**, This Ξ optimized stimulus discrimination. Same style as panel d. **f**, Decision tree classification illustrated. Cascading true/false tests on Ξ elements produce predictions. A random forest uses multiple trees that "vote" on classification.

3.3.2 Dynamical Discrimination Reveals Smaller Changes in Stimulus Characteristics than Firing Rate Substitutes Reveal

Differences between trajectories were sensitively quantified by examining the coefficients of polynomial Ordinary Differential Equations (ODEs) fitted to trajectories by an augmented version of Sparse Identification of Nonlinear Dynamics (SINDy) (Brunton et al., 2016) (see methods, Figure 3.1d). ODEs are represented as a matrix of coefficients, Ξ (Figure 3.2d, 2e). A genetic algorithm decides which coefficients to make nonzero, numerosity varied from 7-20 (median is 12). The ODEs usually captured the central axis and direction of divergence (Figure 3.2d, 2e). To decode (classify) stimulus features, the utility function was the mean cross-validated F_1 score of a random forest classifier (Figure 3.2f) trained on individual Ξ and a regularization penalty (see methods). After the last generation an ensemble of Ξ matrices vote on the classification (see methods). Next overfitting was directly measured by repeating the entire genetic algorithm with random surrogates (Figure 3.3b). If overfitting is minimal (Figure 3a, 3c, but not 3b) the cross-validation performance is a sufficient discriminating statistic. Stimulus feature classification final results (Figure 3.3d) use additional 20-fold holdout out-of-sample generalization. Details of our algorithm, stability analysis of ODEs, and hyperparameters optimization, including choice of: regularization, epoch, dimensions, derivative order, and more is in the appendix A.2. When optimizing for classification (decoding) we call it "dynamical discrimination", when optimizing for ODE goodness of fit (see methods) we call it "best-fit Ξ based discrimination".

Our dynamical discrimination results differed by stimulus-recording category (see above): Some characteristics were better encoded in Ξ than others and excitatory and inhibitory current were different from each other and membrane potential. Random chance would yield correct classification rates (CCR) of 1/8=0.125 for classifying orientation and 1/6≈0.167 for size or contrast. For the set as a whole, dynamical discrimination did statistically significantly greater than chance (median CCR: CCR_{med}=0.288, r_{sdf}=0.239, p=5.5x10^{-27}), according to a one-tailed Wilcoxon signed-rank test, where r_{sdf} is the simple-difference effect size and p is the probability value (see Methods). Individually the following categories were also greater than chance with the criterion p<0.05: $^{L,O}I$ (CCR_{med}=0.144, r_{sdf}=0.243, p=0.0313), $^{E,O}I$ (CCR_{med}=0.256, r_{sdf}=0.20, p=0.0156), $^{L,C}I$ (CCR_{med}=0.292, r_{sdf}=0.258, p=3.05x10^{-5}), $^{E,C}I$ (CCR_{med}=0.275, r_{sdf}=0.258, p=3.05x10^{-5}), $^{R,C}V$ (CCR_{med}=0.275, r_{sdf}=0.224, p=0.0225) and $^{K,C}V$ (CCR_{med}=0.25, r_{sdf}=0.238, p=0.0098), $^{L,S}I$ (CCR_{med}=0.363, r_{sdf}=0.23, p=7.63x10^{-6}), and $^{E,S}I$ (CCR_{med}=0.283, r_{sdf}=0.257, p=3.81x10^{-6}). The only data categories which did not meet either criterion for distinguishability are $^{R,S}V$, (CCR_{med}=0.3), and $^{K,S}V$, (CCR_{med}=0.408), despite having the largest group median there were only five data points and too much variability to draw a conclusion. By pooling results from the categories $^{L,O}I$ and $^{E,O}I$ we reinforced our confidence that fine changes in orientation can be distinguished (CCR_{med}=0.216, r_{sdf}=0.218, p=4.9x10^{-4}).

Since any quality machine learning algorithm should permit better than chance classification we needed to contrast dynamical discrimination with methods that either ignore fluctuations or are agnostic to a dynamical systems interpretation. For simplicity we call all measurements of difference from baseline “deflection” and we used only the

absolute value of this difference. To calculate deflection any spikes were always removed. Then, a baseline unique to each recording (see methods) was subtracted from the average value of a 166 ms window containing the peak of the average response across all trials of all stimuli. This is comparable to the methods of the paper the data was collected for (Adesnik, 2017) and tested to maximize deflection based discrimination. Thus we can compare the results of dynamical discrimination to a random forest classifier trained on evoked deflection, which is a typical firing-rate substitute that ignores fluctuations. We also compared dynamical discrimination to best-fit Ξ based discrimination.

We found that all three methods, deflection based discrimination (Figure 3.3a), best-fit Ξ based discrimination (Figure 3.3c), and dynamical discrimination (Figure 3d), performed better than chance when pooling all data from all categories (see also appendix A.1). However, best-fit Ξ allowed higher CCR than deflection (Figure 3.3c), (CCR_{med}=0.267, 0.228 respectively), a significant difference (r_{sdf}=0.178, p=1.87x10^{-4}). A one-tailed Wilcoxon signed-rank test establishes comparative performance. Best-fit Ξ based discrimination also outperformed deflection based discrimination in a few individual categories while deflection based discrimination was never better (see appendix A.1 table A.1). Finally, dynamical discrimination yielded the highest CCR for pooled and most individual data (best-fit Ξ comparison: r_{sdf}=0.177, p=7.8x10^{-5}, and deflection comparison: r_{sdf}=0.202, p=3.60x10^{-8}) and no method outperformed it in individual categories (see appendix A.1 table A.1). These differences are significant, especially regarding the ability to classify orientation, neurons do receive more fine-grained information than they are able to pass on through action potentials.

One key difference between best-fit Ξ based discrimination (Figure 3.3c) and dynamical discrimination (Figure 3.3d) is that best-fit Ξ failed to show greater than chance modulation of inhibitory activity trajectories (CCR_{med}=0.1256, r_{sdf}=0.1619, p=0.344), but still showed it for excitatory trajectories (CCR_{med}=0.1908, r_{sdf}=0.2381, p=0.039). Thus the extra fidelity from an algorithm that priortizes stimulus discrimination over ODE fitting accuracy was required to find orientation information in inhibitory activity trajectories. For more granular analysis of dynamical discrimination comparisons and validation that deflection can still reliably discriminate between just the most and least preferred stimului see appendix A.1. Performance with different dimensions, types of ODEs, and different epochs of visual stimulation are included in appendix A.2.

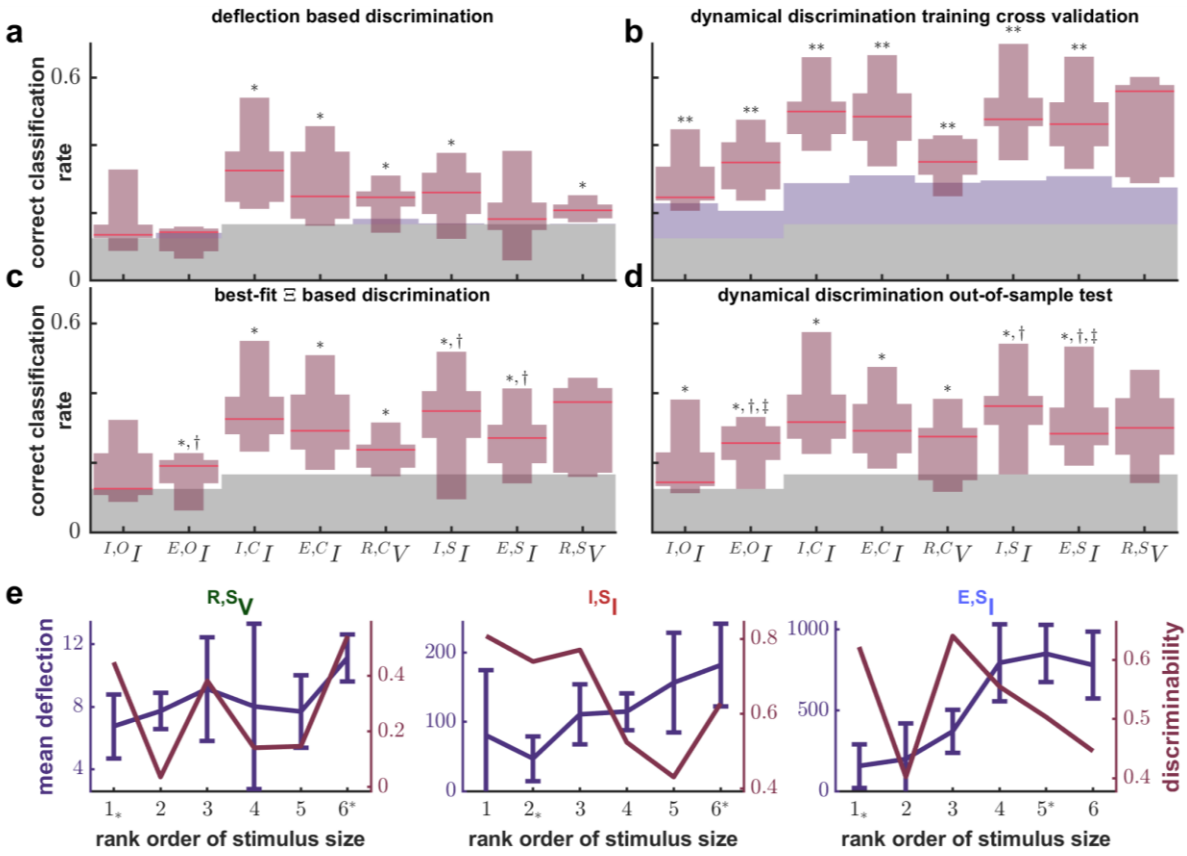

Figure 3.3 | Discrimination performance of multiple methods across categories and compared with deflection for individual cells. **a**, Red shading shows box and whisker plots of correct classification rates (CCR) based on deflection, with red lines indicating medians. Gray shading shows chance level. Columns separate all data by categories containing the same recording type and stimulus characteristic. * indicates significantly greater than chance performance in that category. Discriminating orientation and size ($^{E,S}I$) fails with deflection. **b**, The last generation cross-validation performance of Ξ matrices chosen to optimize discrimination (out-of-sample generalization is panel d). Lavender shading shows overfitting effect (random surrogate median CCR) and is barely visible in panels a and c. All categories (even surrogates) are distinguished from chance. ** indicates distinguishability from random surrogates. **c**, Overfitting vanishes while permitting better than deflection discrimination when Ξ is chosen to optimize ODE modelling. Distinguishability from chance and deflection are indicated with * and † respectively. **d**, Dynamical discrimination retains superiority after out-of-sample testing. Distinguishability from chance, deflection, and best fit results are indicated with *, †, and ‡ respectively. Orientation and size (^{LS}I, $^{E,S}I$) can be discriminated better than with deflection. **e**, Average deflection (purple, error bars indicate standard deviation) and discriminability (dark red, F_1 score of dynamical discrimination) as a function of drifting grating property. The ordinal label marks the abscissa. The least and most preferred stimuli are indicated with ⋆ and * respectively. Left: Membrane potential with spikes removed (mV), Center: inhibitory transmembrane current (pA). Right excitatory transmembrane current (pA).

3.3.3 Dynamical Discrimination is Linked to Distinct and Reliable Presynaptic Population Dynamics

The most and least preferred stimuli for a neuron of interest evoke the largest or smallest (respectively) deflections on average for that neuron. A plot of average deflection vs stimulus characteristic is called a "tuning curve". In Figure 3.3e we use the F_1 score as a measure of discriminability and compare it with the tuning curves obtained for each cell. Neither the least nor the most preferred stimulus coincides with the largest standard deviation of deflection or poorest discriminability. In Figure 3.4a we show that, for every signal source and stimulus type, the most and least preferred stimulus tends to be the most discriminable.

We examined the reliability of a cell's responses to each stimulus type. Defined as the mean deflection divided by the standard deviation, reliability was greater for deflections evoked by the most and least preferred stimuli (Figure 3.4b). To quantify the U-shape of Figure

3.4a and 4b we computed the correlation between reliability and the ordinal-number distance to either the least preferred stimuli or the most (whichever is smallest, see methods). We calculated the correlation between reliability and distance to most/least preferred stimulus to be 0.60 (Pearson correlation coefficient) with p value $p=3.45x10^{-67}$. The same distance value correlated with F_1 score of dynamical discrimination according to 0.26 (Pearson correlation coefficient) with $p=5.04x10^{-12}$. We know that population dynamics evoked by the most and least preferred stimulus are distinct from each other because the average deflection is so different (see also Fig S1b). Because reliability is much higher for these stimuli we know they evoke population dynamics distinct from the intermediate stimuli.

The question becomes whether this reliability pattern is best captured by dynamical system regimes or by random process variability. Linear stability analysis (Strogatz, 2018) defines dynamical systems regimes through analytical properties of real-valued zero-gradient solutions to the ODE, called "fixed-points" (see methods). We found that the most preferred stimulus stood out. Stimuli dissimilar to the preferred stimuli produced a dissimilar number of fixed-points (see appendix A.3, Figure A.7). Additionally, the fraction of net-convergent fixed-points inversely correlated with dissimilarity to the preferred stimulus (Figure A.7). Despite the intentional crudeness of our ODEs (to exaggerate differences) they may further evince distinct stimulus tuned population dynamics. This complements the observation of heightened discriminability for the most and least preferred stimuli.

Crucially, we examined single-trial variability and found increased reliability is not effected through smaller fluctuations or less noise. We defined fluctuation size as the coefficient of variation for single recordings and defined noise as the mean of fractional response residuals (see methods). Fluctuation size and noise both simply decreased with increasing average deflection and did not exhibit a U-shaped trend, suggesting neither are responsible for increased response reliability, alone or in combination (see appendix A.3, Figure A.8). The three observations of dynamical discrimination performance, fluctuation size and noise, and linear stability analysis combined suggest that reliability is rooted in increasingly coherent dynamical patterns.

We then devised an even stronger test of the link to dynamic stability. As shown in Figure 3.4c a plot of reliability versus deflection roughly separates stimuli when labeled according to each cell's preferences. If reliability and average deflection reflect population dynamics unique to each stimulus feature then we expect Ξ matrices will have the same information, but on a single trial basis. Therefore we should be able to further reduce Ξ matrices to just these two variables and use them to predict stimulus features with minimal performance impact.

We re-analysed ensembles of Ξ matrices with random forest regression (see methods) to estimate reliability and average deflection (Figure 3.4d) then fed the estimates to a third random forest classifier to predict the stimulus labels (Figure 4e). While there was a small decrease in performance overall (CCR_{med}=0.283, r_{SDF}=0.141, p=0.0061) this was the second-best method of classification, retaining most of the key results such as better than

chance classification of $_{E,O}I$ and better than deflection performance in $_{E,O}I$, $_{I,S}I$, $_{E,S}I$, and even $_{I,C}I$,. This regression-then-classification approach demonstrates the extraction of latent variables (reliability and average deflection factors) from single trials and relates them to a controlled characteristic (stimulus label) (Whiteway and Butts, 2019). The success of using reliability and deflection like putative state variables adds to the evidence that most and least preferred stimuli are associated with distinct dynamics among the presynaptic populations which drive our observed membrane dynamics.

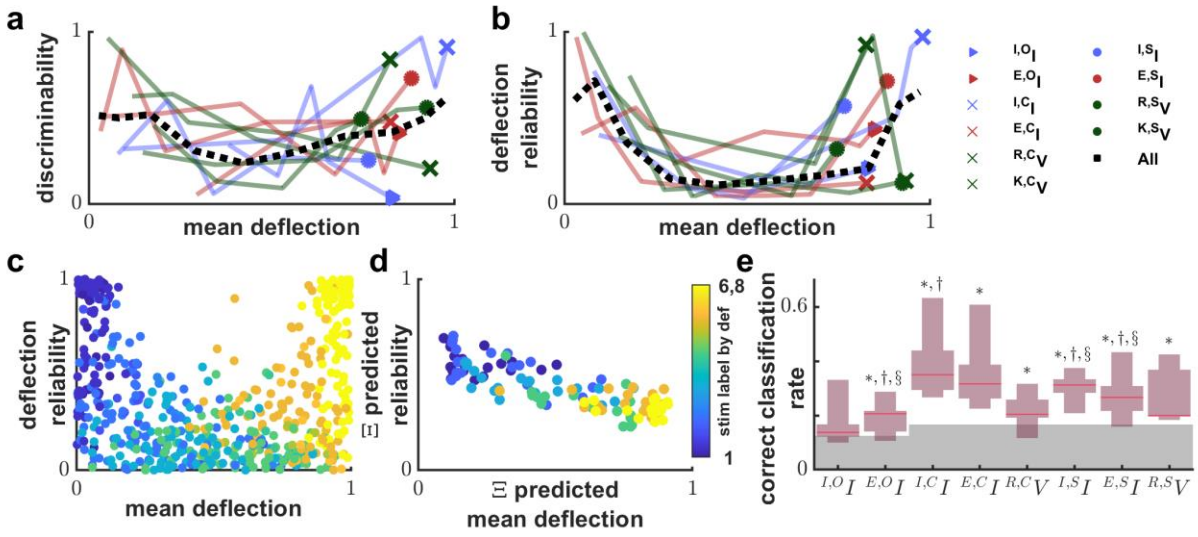

Figure 3.4 | Single trial discriminability depends on dynamical states associated with stimulus selectivity. a, The U shape of normalized discriminability vs normalized mean deflection indicates least and most preferred stimuli are more discriminable. For each decile of normalized mean evoked deflection it's median is plotted against the median normalized discriminability (F_1 score of dynamical discrimination). The legend right of panel b maps color and end-point marker to data categories. Dashed line indicates pooling of all categories. **b**, Deflection reliability (inverse coefficient of variation) shows a more prominent U trend (same style as a). **c**, Stimuli are roughly separable on a scatter plot of reliability and mean deflection across all data points. Stimuli are re-labeled and colored by their rank of average evoked deflection (color bar is right of panel d) **d**, A plot of trial-by-trial predictions of cross-trial means. Predictions are from random forest regressors trained on Ξ matrices re-appropriated from dynamical discrimination. What separability remains now extends to a trial-by-trial basis. **e**, Same as in Figure 3.3d except now the context (original stimulus label) is inferred using only the predicted state variables (reliability and deflection) from panel d. Distinguishability from chance is denoted with *, from deflection with †, and § indicates significantly worse performance than Figure 3.3d (the reverse was never true). The key results from Figure 3.3d are reproduced despite stripping information down to just two understandable variables: reliability and

deflection. Furthermore ^{LC}I was distinguishable from deflection which did not occur for Figure 3.3d.

3.3.4 Dynamical Discrimination is Corroborated by Biologically Plausible Modeling

Because dynamical discrimination is based on estimating the coefficients of ODEs, it is incumbent on us to test it with the coefficients of known ODEs. Our overall strategy, illustrated in Figure 3.5, is to challenge dynamical discrimination by using resampling and neuron models to degrade data generated from a known system. Instead of using Ξ matrices to predict a label from a small set of possible labels, the genetic algorithm used regression to accurately estimate small changes to a parameter of the Lorenz system, ρ, as we varied it between the integers 20 to 40.

The Lorenz system is:

$$\begin{aligned} \frac{dX}{dt} &= \sigma(Y - X) \\ \frac{dY}{dt} &= X\,(\rho - Z) - Y \\ \frac{dZ}{dt} &= XY - \beta Z \end{aligned} \qquad (3.1)$$

We chose ρ to include a Hopf bifurcation at $\rho = \frac{\sigma(\sigma+\beta+3)}{(\sigma-\beta-1)} \approx 24.7$and to explore chaotic regimes. We used the X dimension of the Lorenz system but resampled each trial $X'(t) = X(t/\tau)$where τ is chosen such that 90% of the signal power was in Fourier modes below 300 Hz for each trial (see methods, Fig 5b). As seen in Figure 3.5d our approach excels at

predicting the value of ρ from a sample of time series data. Median absolute percent error between the predicted value of ρ and the real value has a median of 1.79% with a 25th percentile, P_{25}= 0.87% and a 75th percentile P_{75}=3.46%, verifying the identification of ρ for nearly ideal conditions.

We followed up by distorting $X'(t)$with successively more complex neuron-like transformation. We fed the resampled X dimension from the Lorenz equation into a single-compartment neuron model by shifting and rescaling to be consistent with injected current (no synapses) on the order of nano amps $I(t) = \alpha_1(X'(t) - \zeta_1)$ where ζ_1 and α_1 are chosen to produce the desired amount of spiking (see methods, Figure 3.5b, 5c). Median error was 5.45% with P_{25}=2.52% and P_{75}=9.18% (Figure 3.5d). All values reported in the text come from data where spikes were removed (see methods) which made little difference for any model. The still good performance shows that transformation by membrane dynamics does little to interfere with our methods.

In a biologically plausible attractor network, the dynamics actually govern a point process with inputs to a neuron arriving as discrete events. Therefore we fed the single resampled dimension, X from the Lorenz equation into a morphologically complex multi-compartment neuron model (Palmer and Stuart, 2009) by shifting and rescaling it to be consistent with the instantaneous event probability of an inhomogeneous Poisson process $P(t) = \alpha_2(X'(t) - \zeta_2)$ where ζ_2 and α_2 are chosen to produce the desired amount of spiking (see methods, Figure 3.5b). Models with fewer (excitatory only) synapses results in a membrane potential time series that has large and sporadic synaptic potentials (impulses).

Initially adding more synapses produces a signal that better visually approximates the input time series, but eventually increases dynamical complexity and spiking unpredictability (Fig 5c). This more complex type of model yielded a median error of 12.02% with P_{25}=8.11% and P_{75}=19.99% (Figure 3.5d). Thus synaptic type distortions have degraded ODE fitting approaches but not left them unworkable.

To directly compare with experimental dynamical discrimination results we treated the ρ values [22,25,28,31,34,37] as if they were distinct stimuli labeled one through six and re-analyzed the corresponding Ξ matrices by training classifiers to predict the labels. Chance level is $1/6 \approx 0.167$. We also categorize model trials according to the number of spikes and synapses because they may be confounding factors. Single-compartment neuron data formed 12 categories according to similar spiking levels (see methods). Multi-compartment models yielded 18 spiking categories (limited to models with 245 to 290 synapses for Figure 3.5e) and 35 synapse numerosity categories (limited to trials with four or fewer spikes for Figure 3.5e). The impact of spike rate and synapse numerosity is shown in Figure 3.5e. For single-compartment data (spikes removed) median CCR for the 12 categories ranged from 0.563 to 0.778. The median CCR was indifferent to the level of spiking. By comparison, median CCR was low and inconsistent for the 18 multi-compartment spiking categories, with values ranging from 0.05 to 0.5. Nonetheless, no association with spiking passed statistical muster. However, synapse numerosity categories did show a trend. Median CCR values ranged between 0.05 and 0.393. CCR increased quickly for increasing numbers of synapses, peaking at around 162 synapses, then declined steadily. By comparing Figure 3.5e to Figure 3.3d we can see that the range of

CCR values overlaps with the median CCR values for in-vivo dynamical discrimination. Median CCR from Figure 3.3d ranged from 0.144 to 0.363 with an overall median of 0.283. Thus we see our results for in-vivo data are near the ceiling for this implementation of dynamical discrimination and in line with expectations for synaptic impulses that are completely governed by a dynamical system being manipulated through a bifurcation and chaos.

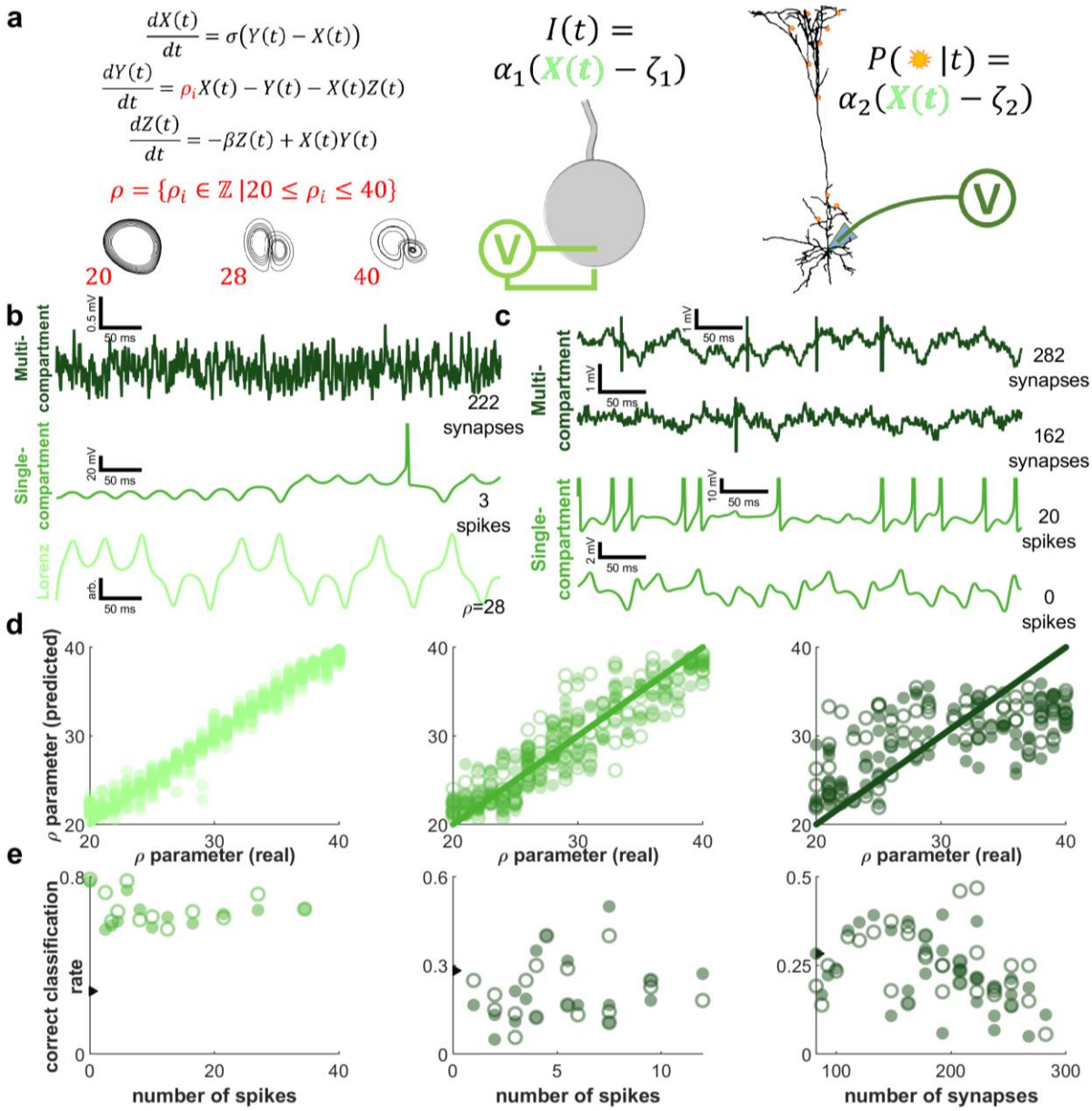

Figure 3.5 | Modeling tests confounding factors for dynamical discrimination and matches experimental results. a, Our modeling paradigm, illustrated. Left: We tested regressing Ξ onto integer values of the Lorenz system parameter ρ from 20 to 40 (spanning a Hopf bifurcation and chaos). Center: a single-compartment neuron (X as current injection). Right: a multi-compartment model with dendritic spines (NEURON shape plot, X governs synapse transmission probability). **b,** Simulated membrane potential. Top: multi-compartment model (dark green). Middle: single-compartment neuron (medium green). Bottom: X from Lorenz system (light green). **c,** Examples of possible confounds. Top: two traces showing differences associated with synapse numerosity. Bottom: two traces showing action potentials dominating Lorenz dynamics. **d,** Regressions of Ξ onto ρ plotted against true ρ values. Left: Fitting Ξ to X performed well (light green). Center: Fitting Ξ to membrane potential of single-compartment neurons moderately reduced accuracy (medium green). Right: Multi-compartment models significantly degraded regressions (dark green). Performance is similar whether spikes are removed (filled) or not (open). **e,** Correct classification rate (CCR) vs possible confounds. Ξ matrices trained for regression are reappropriated for classification and ρ is limited to [22,25,28,32,35,38] (chance 1/6≈0.167). Left: Dynamical discrimination is robust to

spiking for single-compartment models (medium green). Center: Spiking also has limited impact for multi-compartment models (dark green). Right: Synapse numerosity (model complexity) is impactful (dark green). Arrow ticks indicating median CCR for data in Figure 3.3d evince modeling and experiment agreement but higher potential for dynamical discrimination with continuous dynamics.

3.4 Discussion

This work presented compelling observations based on a novel hypothesis-dependent machine learning (Bishop, 2013) algorithm. First, membrane potential and transmembrane current recorded from neurons in mouse primary visual cortex underwent dimensionality expansion. This yielded trajectories of neural activity which appear to be stimulus modulated. Second, motivated by attractor network principles (Amit and Amit, 1992; Goldberg et al., 2004; Eliasmith, 2005; Wu and Amari, 2005; Miconi et al., 2016; Chambers and Rumpel, 2017), ordinary differential equation models (Ξ matrices) were fitted to individual trajectories. Ξ matrices compactly describe each trajectory and were used to predict what stimulus presentation coincided with each trajectory. Called dynamical discrimination, this algorithm more accurately predicted fine changes in orientation, contrast, and size of drifting gratings than predictions made from firing rate substitutes (deflection) and alternatives. Furthermore, only dynamical discrimination had the fidelity to confirm findings about orientation selectivity differences between excitatory and inhibitory synaptic mechanisms. Thus dynamical discrimination is an advanced tool for sensing population dynamics among the neurons inputting to a neuron subjected to intracellular recording. Third, stimuli evoking extremes of average deflection also evoked the most reliable deflections but not the smallest fluctuations, or least noise. Dynamical discrimination matched reliability patterns, and further analysis of Ξ matrices show that

the least and most preferred stimuli evoke distinctive and coherent dynamical regimes. If the concept of binary feature detectors (Keller and Mrsic-Flogel, 2018) and attractor network principles apply then these may be like task-positive and task-negative dynamical regimes (Churchland et al., 2010; Gallego et al., 2018) such that intermediate stimuli evoke less reliable deflections because they either alternately evoke these two regimes or evoke inherently less coherent dynamical regimes in the presynaptic population. Lastly, modeling validated the level of accuracy. We showed that dynamical discrimination excels with continuous nonlinear transformations of underlying dynamics, but transforming dynamics into a point process (like synaptic transmission) degrades performance to the experimentally observed levels. These four results show that: i) Dynamical discrimination is a powerful method for time series and trajectory classification. ii) Attractor network principles can be applied to primary visual cortex and to single neuron recordings, and iii) dimensionality expansion and dynamical discrimination lets researchers patch into upstream network by intracellularly recording from single neurons.

The dimensionality expansion of intracellular recordings in this work (Figure 3.2) counterpoints the commonly used dimensionality reduction of population activity (Cunningham and Yu, 2014). Population recording methods have improved in recent years (Greenberg et al., 2008; Viventi et al., 2011; Steinmetz et al., 2018), likewise for efforts to isolate and group single units by functional and anatomical relevance (Palmer and Stuart, 2009; Pachitariu et al., 2016; Bassett and Sporns, 2017). Subsequent analysis usually includes dimensionality reduction. In contrast, dimensionality expansion on single neurons instead may yield similar insights about neuronal manifolds (Gallego et al., 2018) but the

grouping is naturally defined as: all presynaptic neurons (Kuhn et al., 2003; Ko et al., 2011; Yaşar et al., 2016). The general principle of dimensionality expansion is familiar to neuroscience (e.g. separating a signal into time-varying oscillatory modes) though not usually for reconstructing trajectories.

Researchers have tried many ways to map synaptic activity recorded at the soma to population events and dynamics (Jagadeesh et al., 1992; Baudot et al., 2013; Perks and Gentner, 2015; Yaşar et al., 2016; Wright et al., 2017). Machine learning algorithms are sensible options but often don't permit scientific inference beyond their predictions themselves (Bzdok and Yeo, 2017; Whiteway and Butts, 2019)·. Hypothesis-dependent methods (Bishop, 2013; Butner et al., 2019) leverage computational assets to fit a model to data and exploit the model for a new purpose (Bishop, 2013; Daniels and Nemenman, 2015; Brunton et al., 2016; Mangan et al., 2016; Kaiser et al., 2018). If the model is fitted poorly or not applicable the algorithm performs poorly, thereby testing the hypothesis motivating the choice of model. This is why important earlier works (Lainscsek and Sejnowski, 2015; Oprisan et al., 2015; Brunton et al., 2016) inspired us to develop dynamical discrimination.

Dynamical discrimination's power to identify latent factors (Gallego et al., 2018; Pandarinath et al., 2018; Whiteway and Butts, 2019) (i.e. context) may advance traditional single-neuron topics like orientation tuning (Seriès et al., 2004; Li et al., 2012) and interplay between excitatory and inhibitory populations (Kuhn et al., 2003; Adesnik, 2017) . Pairing dynamical discrimination on intracellular recordings with comparisons across

brain regions, multiple recording and/or stimulation sites, or cell staining and tracing promises new insights into functional specificity and connectivity (Mangan et al., 2016; Bassett and Sporns, 2017) because intracellular recordings observe in-situ network propagation at a natural bottleneck.

For experimental data the dynamical discrimination algorithm had limited classification accuracy. However, this is expected and exceeds all compared methods. Modelling showed that synaptic transmissions completely governed by a simple dynamical system produced these levels of dynamical discriminability, while validating high accuracy in more ideal scenarios. Dynamical discrimination was further corroborated by recapitulating historical findings such as differential tuning among excitatory and inhibitory populations (Priebe and Ferster, 2005; Adesnik, 2017), and differences in discriminability between early and late epochs (Müller et al., 2001; Wang et al., 2010; Clawson et al., 2017) (see appendix A.2). Furthermore, we gained new information regarding patterns of response reliability (Churchland et al., 2010). Performance may improve since this was a simplistic version of the algorithm. Single cells yielded too few recordings for deep learning alternatives, but fitting ODE models reduces the trajectories to fewer characteristics than training points, thereby avoiding overfitting. Thus we conclude that limited accuracy is expected for this first version of dynamical discrimination and belies its power as an analytical tool.

This work advances physics and neuroscience in several ways. We must specify and isolate relevant parts of the brain before we can elucidate their interactions. The brain has natural partitions such as layers, nuclei, and cortical columns, but imaging fields and electrode

arrays often overlap or partially cover them. An elementary natural brain partition is the neurons synapsing onto a neuron of interest. We applied dimensionality expansion to intracellular recordings, yielding trajectories related to those from dimensionality reduction on hypothetical high-dimensional recordings of these neurons. We used a dynamical discrimination algorithm applicable to trajectories of any provenance. It was easier to detect orientation sensitively modulating excitatory activity trajectories than inhibitory activity trajectories. Excitatory thalamic projections bring orientation information to V1 (Sun et al., 2016), but deflection misses fine distinctions. Thus we observed detailed dynamics from a natural partition in the brain. Because this works on a trial-by-trial basis (Perks and Gentner, 2015) not an average over trials, we have a versatile tool for investigating neural representation. Our machine learning approach incorporates a hypothetical paradigm (attractor networks) into its core apparatus. Consequently it demonstrated utility beyond classification by connecting stimulus tuning to attractor network principles. Because of these demonstrations, machine learning has a future in science beyond the black-box, decades of single-neuron intracellular recordings can be re-analyzed for population insights, and the attractor network paradigm has come to primary visual cortex.

References

Adesnik H (2017) Synaptic mechanisms of feature coding in the visual cortex of awake mice. Neuron 95:1147–1159.e4.

Adesnik H, Bruns W, Taniguchi H, Huang ZJ, Scanziani M (2012) A neural circuit for spatial summation in visual cortex. Nature 490:226–231.

Amit DJ, Amit DJ (1992) Modeling Brain Function: The World of Attractor Neural Networks. Cambridge University Press.

Anon (n.d.) ModelDB: Membrane potential changes in dendritic spines during APs and synaptic input (Palmer & Stuart 2009). Available at: https://senselab.med.yale.edu/modeldb/ShowModel?model=120798 [Accessed January 28, 2020].

Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. Stat Surv 4:40–79.

Ayaz A, Saleem AB, Schölvinck ML, Carandini M (2013) Locomotion controls spatial integration in mouse visual cortex. Curr Biol 23:890–894.

Bassett DS, Sporns O (2017) Network neuroscience. Nat Neurosci 20:353–364.

Baudot P, Levy M, Marre O, Monier C, Pananceau M, Frégnac Y (2013) Animation of natural scene by virtual eye-movements evokes high precision and low noise in V1 neurons. Front Neural Circuits 7:206.

Bender R, Lange S (2001) Adjusting for multiple testing—when and how? J Clin Epidemiol 54:343–349.

Bennett C, Arroyo S, Hestrin S (2013) Subthreshold mechanisms underlying state-dependent modulation of visual responses. Neuron 80:350–357.

Berens P, Ecker AS, Cotton RJ, Ma WJ, Bethge M, Tolias AS (2012) A fast and simple population code for orientation in primate V1. J Neurosci 32:10618–10626.

Bishop CM (2013) Model-based machine learning. Philos Trans A Math Phys Eng Sci 371:20120222.

Bondanelli G, Ostojic S (2018) Coding with transient trajectories in recurrent neural networks. arXiv [q-bioNC] Available at: http://arxiv.org/abs/1811.07592.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Breiman L (2001) Random Forests. Mach Learn 45:5–32.

Brunton SL, Brunton BW, Proctor JL, Kaiser E, Kutz JN (2017) Chaos as an intermittently forced linear system. Nat Commun 8:19.

Brunton SL, Proctor JL, Kutz JN (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc Natl Acad Sci U S A 113:3932–3937.

Butner JE, Munion AK, Baucom BRW, Wong A (2019) Ghost hunting in the nonlinear dynamic machine. PLoS One 14:e0226572.

Bzdok D, Yeo BTT (2017) Inference in the age of big data: Future perspectives on neuroscience. Neuroimage 155:549–564.

Chambers AR, Rumpel S (2017) A stable brain from unstable components: Emerging concepts and implications for neural computation. Neuroscience 357:172–184.

Churchland MM et al. (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nat Neurosci 13:369–378.

Clawson WP, Wright NC, Wessel R, Shew WL (2017) Adaptation towards scale-free dynamics improves cortical stimulus discrimination at the cost of reduced detection. PLoS Comput Biol 13:e1005574.

Clerc M, Coullet P, Tirapegui E (2001) The stationary instability in quasi-reversible systems and the Lorenz pendulum. Int J Bifurcat Chaos 11:591–603.

Cunningham JP, Yu BM (2014) Dimensionality reduction for large-scale neural recordings. Nat Neurosci 17:1500–1509.

Daelli V, Treves A (2010) Neural attractor dynamics in object recognition. Exp Brain Res 203:241–248.

Daniels BC, Nemenman I (2015) Automated adaptive inference of phenomenological dynamical models. Nat Commun 6:8133.

Duysens J, Schaafsma SJ, Orban GA (1996) Cortical off response tuning for stimulus duration. Vision Res 36:3243–3251.

Eliasmith C (2005) A unified approach to building and controlling spiking attractor networks. Neural Comput 17:1276–1314.

Fu Y, Tucciarone JM, Espinosa JS, Sheng N, Darcy DP, Nicoll RA, Huang ZJ, Stryker MP (2014) A cortical circuit for gain control by behavioral state. Cell 156:1139–1152.

Gallego JA, Perich MG, Naufel SN, Ethier C, Solla SA, Miller LE (2018) Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. Nat Commun 9:4233.

Gao E, DeAngelis GC, Burkhalter A (2010) Parallel input channels to mouse primary visual cortex. J Neurosci 30:5912–5926.

Goldberg JA, Rokni U, Sompolinsky H (2004) Patterns of ongoing activity and the functional architecture of the primary visual cortex. Neuron 42:489–500.

Greenberg DS, Houweling AR, Kerr JND (2008) Population imaging of ongoing neuronal activity in the visual cortex of awake rats. Nat Neurosci 11:749–751.

Gutnisky DA, Dragoi V (2008) Adaptive coding of visual information in neural populations. Nature 452:220–224.

Harriman S, Patel J (2014) Text recycling: acceptable or misconduct? BMC Med 12:148.

Hastie T, Tibshirani R, Friedman J (2009a) Linear Methods for Regression. In: The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Hastie T, Tibshirani R, Friedman J, eds), pp 43–99. New York, NY: Springer New York.

Hastie T, Tibshirani R, Friedman J (2009b) Ensemble Learning. In: The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Hastie T, Tibshirani R, Friedman J, eds), pp 605–624. New York, NY: Springer New York.

Jagadeesh B, Gray CM, Ferster D (1992) Visually evoked oscillations of membrane potential in cells of cat visual cortex. Science 257:552–554.

Johnson JK, Wright NC, Xià J, Wessel R (2019) Single-Cell Membrane Potential Fluctuations Evince Network Scale-Freeness and Quasicriticality. J Neurosci 39:4738–4759.

Jouhanneau J-S, Poulet JFA (2019) Multiple Two-Photon Targeted Whole-Cell Patch-Clamp Recordings From Monosynaptically Connected Neurons in vivo. Frontiers in Synaptic Neuroscience 11 Available at: http://dx.doi.org/10.3389/fnsyn.2019.00015.

Kaiser E, Kutz JN, Brunton SL (2018) Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. Proc Math Phys Eng Sci 474:20180335.

Keller GB, Mrsic-Flogel TD (2018) Predictive processing: a canonical cortical computation. Neuron 100:424–435.

Kerby DS (2014) The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. Comprehensive Psychology 3:11.IT.3.1.

Kim CK, Adhikari A, Deisseroth K (2017) Integration of optogenetics with complementary methodologies in systems neuroscience. Nat Rev Neurosci 18:222–235.

Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, Mrsic-Flogel TD (2011) Functional specificity of local synaptic connections in neocortical networks. Nature 473:87–91.

Kostelich EJ, Schreiber T (1993) Noise reduction in chaotic time-series data: A survey of common methods. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 48:1752–1763.

Kuhn A, Aertsen A, Rotter S (2003) Higher-order statistics of input ensembles and the response of simple model neurons. Neural Comput 15:67–101.

Lainscsek C, Sejnowski TJ (2015) Delay differential analysis of time series. Neural Comput 27:594–614.

Liang Z, Shen W, Sun C, Shou T (2008) Comparative study on the offset responses of simple cells and complex cells in the primary visual cortex of the cat. Neuroscience 156:365–373.

Liu B-H, Li P, Li Y-T, Sun YJ, Yanagawa Y, Obata K, Zhang LI, Tao HW (2009) Visual receptive field structure of cortical inhibitory neurons revealed by two-photon imaging guided recording. J Neurosci 29:10520–10532.

Li Y-T, Ma W-P, Li L-Y, Ibrahim LA, Wang S-Z, Tao HW (2012) Broadening of inhibitory tuning underlies contrast-dependent sharpening of orientation selectivity in mouse visual cortex. J Neurosci 32:16466–16477.

London M, Häusser M (2005) Dendritic computation. Annu Rev Neurosci 28:503–532.

Lorenz EN (1963) Deterministic Nonperiodic Flow. J Atmos Sci 20:130–141.

Mangan NM, Brunton SL, Proctor JL (2016) Inferring biological networks by sparse identification of nonlinear dynamics. IEEE Transactions on Available at: https://ieeexplore.ieee.org/abstract/document/7809160/.

Maravall M, Petersen RS, Fairhall AL, Arabzadeh E, Diamond ME (2007) Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. PLoS Biol 5:e19.

Mazor O, Laurent G (2005) Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. Neuron 48:661–673.

Miconi T, McKinstry JL, Edelman GM (2016) Spontaneous emergence of fast attractor dynamics in a model of developing primary visual cortex. Nat Commun 7:13208.

Müller JR, Metha AB, Krauskopf J, Lennie P (2001) Information conveyed by onset transients in responses of striate cortical neurons. J Neurosci 21:6978–6990.

Niell CM, Stryker MP (2008) Highly selective receptive fields in mouse visual cortex. J Neurosci 28:7520–7536.

Oprisan SA, Lynn PE, Tompa T, Lavin A (2015) Low-dimensional attractor for neural activity from local field potentials in optogenetic mice. Front Comput Neurosci 9:125.

Pachitariu M, Steinmetz NA, Kadir SN, Carandini M, Harris KD (2016) Fast and accurate spike sorting of high-channel count probes with KiloSort. In: Advances in neural information processing systems, pp 4448–4456. papers.nips.cc.

Pakan JM, Lowe SC, Dylda E, Keemink SW, Currie SP, Coutts CA, Rochefort NL (2016) Behavioral-state modulation of inhibition is context-dependent and cell type specific in mouse visual cortex. Elife 6 Available at: https://doi.org/10.7554/eLife.14985.001.

Palmer LM, Stuart GJ (2009) Membrane potential changes in dendritic spines during action potentials and synaptic input. J Neurosci 29:6897–6903.

Pandarinath C, O'Shea DJ, Collins J, Jozefowicz R, Stavisky SD, Kao JC, Trautmann EM, Kaufman MT, Ryu SI, Hochberg LR, Henderson JM, Shenoy KV, Abbott LF, Sussillo D (2018) Inferring single-trial neural population dynamics using sequential auto-encoders. Nat Methods 15:805–815.

Perks KE, Gentner TQ (2015) Subthreshold membrane responses underlying sparse spiking to natural vocal signals in auditory cortex. Eur J Neurosci 41:725–733.

Phillips WA, Singer W (1974) Function and interaction of on and off transients in vision. I. Psychophysics. Exp Brain Res 19:493–506.

Polack P-O, Friedman J, Golshani P (2013) Cellular mechanisms of brain state–dependent gain modulation in visual cortex. Nat Neurosci 16:1331–1339.

Priebe NJ, Ferster D (2005) Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. Neuron 45:133–145.

Rabinovich MI, Varona P, Selverston AI, Abarbanel HDI (2006) Dynamical principles in neuroscience. Rev Mod Phys 78:1213.

Reimer J, Froudarakis E, Cadwell CR, Yatsenko D, Denfield GH, Tolias AS (2014) Pupil fluctuations track fast switching of cortical states during quiet wakefulness. Neuron 84:355–362.

Seriès P, Latham PE, Pouget A (2004) Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. Nat Neurosci 7:1129–1135.

Singer W, Phillips WA (1974) Function and interaction of on and off transients in vision. II. Neurophysiology. Exp Brain Res 19:507–521.

Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Inf Process Manag 45:427–437.

Solomon SG, Kohn A (2014) Moving sensory adaptation beyond suppressive effects in single neurons. Curr Biol 24:R1012–R1022.

Steinmetz NA, Koch C, Harris KD, Carandini M (2018) Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. Curr Opin Neurobiol 50:92–100.

Strogatz SH (2018) Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. CRC Press.

Sun W, Tan Z, Mensh BD, Ji N (2016) Thalamus provides layer 4 of primary visual cortex with orientation- and direction-tuned inputs. Nat Neurosci 19:308–315.

Vinck M, Batista-Brito R, Knoblich U, Cardin JA (2015) Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. Neuron 86:740–754.

Viventi J et al. (2011) Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity in vivo. Nat Neurosci 14:1599–1605.

Wang Q, Webber RM, Stanley GB (2010) Thalamic synchrony and the adaptive gating of information flow to cortex. Nat Neurosci 13:1534–1541.

Whiteway MR, Butts DA (2019) The quest for interpretable models of neural population activity. Curr Opin Neurobiol 58:86–93.

Wilcoxon F (1992) Individual Comparisons by Ranking Methods. In: Breakthroughs in Statistics: Methodology and Distribution (Kotz S, Johnson NL, eds), pp 196–202. New York, NY: Springer New York.

Wright NC, Hoseini MS, Yasar TB, Wessel R (2017) Coupling of synaptic inputs to local cortical activity differs among neurons and adapts after stimulus onset. J Neurophysiol 118:3345–3359.

Wu S, Amari S-I (2005) Computing with continuous attractors: stability and online aspects. Neural Comput 17:2215–2239.

Yaşar TB, Wright NC, Wessel R (2016) Inferring presynaptic population spiking from single-trial membrane potential recordings. J Neurosci Methods 259:13–21.

Chapter 4: Summary and Future Work

Our findings provide a solid framework for investigating neural code through the whole-cell recordings. In chapter two we carried out a rigorous comparison between the statistics of membrane potential fluctuations and the statistics of neuronal avalanches. This not only solidified evidence for criticality in neural systems, it demonstrated that scale-freeness provides a reason to expect that other properties of neural populations can be studied in single whole cell recordings. What's more, our modeling investigations lead to some general insights on criticality models that may strengthen connections between the observations of criticality and homeostasis (Ma et al., 2019). In our model the criticality condition, critical branching where one action potential triggers one other on average, also corresponded to a critical coarse graining condition where the instantaneous firing rate of each neuron matched the firing rate of its input population. The general properties of critical averaging network models provide fertile ground for future explorations of computational principles, connections to homeostatic mechanisms, and the relationship between single neuron activity and population activity. Most impactfully, critical averaging and scale-freeness motivate investigations of population neural code in single intracellular recordings.

In chapter three we introduced a method rooted in the attractor network paradigm of neural computation (Amit and Amit, 1992) for testing neural code hypotheses using single whole-cell recordings. By using delay-embedding based dimensionality expansion we were

able to fit ODEs to small snippets from single whole-cell recordings by extended existing methods (Brunton et al., 2016). The coefficients of these equations evinced stimulus tuning, thereby providing a basis for neural code rooted in dynamical systems theory. We call this algorithm “dynamical discrimination”. Tested on in-vivo whole cell recordings in mouse primary visual cortex, we were able to discriminate fine distinctions in drifting grating orientation, size, and contrast. The neurons already evinced stimulus tuning based on deflection of recorded signal from baseline, but the distinctions permitted were crude compared to dynamical discrimination. Furthermore, we recapitulated results from prior population recordings that found orientation tuning in excitatory synapses resulting from thalamocortical projections. The correct classification rate was well below deep learning state of the art but more than high enough to draw useful conclusions. Modeling revealed that the algorithm excels in ideal circumstances (directly on continuous dynamical systems). Modeling showed that the lower performance in experiments is typical for signals that approximate continuous dynamics with a point process using a morphologically complex neuron model. Our modeling investigation provides useful targets for continued improvement of dynamical discrimination. What’s more the ODE fits themselves showed consistency across cells despite being intentionally crude and fitted to brief snippets. Thus, future investigations into the ODE models themselves and the trajectory features they capture are merited. More pressing than improving dynamical discrimination itself, is the application of it to diverse situations where the results can be compared to population recording methods. By comparing to population recordings we can gain powerful new insights into the neural code.

4.1 The Implications for Scale-Freeness and Criticality, Neural Code, and Data Analysis in Neuroscience

The work in these chapters was motivated by three complimentary aims. First, we aimed to investigate the mysterious ubiquity of scale-free signals in brain systems (He, 2014). Second, we aimed to gain an insider's perspective on brain function and be able to directly intercept the neural code at a point of convergence and well-defined functionality: a single neuron. Third, we aimed to develop generally useful data analysis software and demonstrate the effective use of machine learning as part of the scientific inference process.

4.1.1 Aim One: Expand and Clarify the Study of Scale-Freeness and Criticality in Neural Systems

These aims are grounded in neuroscience, physics, and the practice of science. To properly motivate the first aim: to study the origins, extent, and implications of scale-freeness; we articulate basic features of neural systems and critical systems and the similarities between them in the following text. From the earliest days of neuroscience, it has been observed that for each neuron there is a group of disparate presynaptic neurons which converge and provide input. The term "neural action" refers to the process by which postsynaptic neurons aggregate the activity flowing from these inputs (Buice and Cowan, 2009). In recent decades evidence has repeatedly emerged that this process is not passive and that relatively complex operation can take place within a neuron's dendritic arborization (London and Häusser, 2005; Brunel et al., 2014; Gidon et al., 2020) . However, complex

actions may still have simple effects. A similar process is known to physicists as coarse graining and rescaling: a system, often nodes on a lattice, is partitioned and each partition undergoes a rescaling operation. This produces a new smaller system which is distinct from the first. In the simplest form a coarse-graining and rescaling operation simply computes the average of neighboring lattice sites, like resizing a digital imaging. These operators can be much more complex, but their effect is the same: to produce a more "concentrated" version of the initial system. Coarse graining and rescaling is usually applied to systems that have phase transitions. A phase transition is change in the gross physical properties (e.g the water ice phase transition or ferromagnetic phase transition). For systems at the critical point of a phase transition there exists some coarse graining and rescaling operator that produces statistically identical daughter systems. Thus, the critical point is said to be a "fixed point" of the coarse-graining-and-rescaling operation. In cases like that, when the coarse-graining-and-rescaling operation is relevant to critical phenomena, the operation is known as a renormalization group (RG) operator. When a researcher has found an RG operator and applies it to a system that is not at a critical point each daughter system becomes a more extreme version of the parent system. In ferromagnetic systems this means that the apparent temperature of the new "concentrated" system gets closer to zero or infinity. This concentration makes the original phase more and more apparent if the RG operation is iterated. Thus, iterating an RG operator provides a way to classify systems. These two concepts, neural action and coarse-graining-and-rescaling bear a superficial resemblance that has only recently been noticed, expounded upon, and implemented by others (Mehta and Schwab, 2014; Koch-Janusz and Ringel, 2018; Li and Wang, 2018). By hunting for neural code buried in the results of neural

action we can better characterize it, and further elucidate the computational roles of neural action. As such an investigation proceeds, similarities between RG operators and neural action will either become more apparent or be eliminated. In any case the details of neural action suggest that whole-cell recordings can be rich sources of information about the neural code.

The historical goal of studying criticality in neural systems is to test the possibility that it can explain widespread observations of scale-freeness (Beggs and Plenz, 2003). Often this concept is captured by power-law statistics, where a change of scale (i.e. coarse graining and rescaling) does not modify the functional form of a probability distribution. A counter example would be a model network where the probability of a connection between two neurons decays exponentially with the distance between them. The rate of decay is controlled by a characteristic length in the exponent, thus if an analyst decides to work in different units (e.g. mm instead of um) the value of the decay rate must be updated. If connection probability were power-law distributed the rate is controlled by a unitless exponent. Therefore, the decay rate depends on the ratio of lengths and is not affected by a change of variables. When activity propagates in a scale-free manner any event of any size has a non-zero probability to propagate to all parts of the system. In critical systems, but not scale-free systems more generally, spurts of activity of any fixed size or duration will have the same average temporal profile (Sethna et al., 2001). Thus, some of the dynamical details manifest at larger scales should persist at the smallest scales of observation and be recoverable. We cannot expect this for every kind of dynamical property in general. Accounting for features that can and cannot be recovered from single cell observations

should be informative about the nature of scale-freeness in brain systems and highlight an effective route for testing the criticality hypothesis.

4.1.2 Aim Two: Articulate Intrinsic Representations of Neural Dynamics and Neural Code

Our second aim is to gain an insider's perspective on brain function and be able to directly intercept the neural code at a point of convergence and well-defined intrinsic functionality: a single neuron. Well-defined means that few things essential to the function are un-observed, and the brain reacts to the observations as the experimenter has observed them. Intrinsic means that the function does not depend on any actions or assumptions taken by any researcher. For population recording methods identifying the neural code often begins with an enumeration of the activity of a large list of individual neurons. It is unlikely that this can be reconstructed as the basis for neural code using whole-cell patch-clamp techniques at the soma. For this reason, we look to the attractor network paradigm for a neural code basis. In order to understand the impact and significance of our freely available algorithm for assess attractor network theories of neural code, it is helpful to recount some rudiments of the attractor paradigm and how it is currently interpreted.

The network attractor paradigm arose in the early days of neural network simulation. Even today a common variety of artificial neural network classifier functions by defining an output layer and is structured such that all inputs activates only one neuron in this output layer. Each of these output neurons corresponds to one of the possible classification outcomes. Thus, as an input is fed to the network and activity is allowed to propagate

through the layers it is “attracted” to one neuron or the other. This is a rudimentary attractor network, one which always converges to a point in the same number of steps (Hertz, 1995; Eliasmith, 2007). In so far as it models real nervous systems internal layers analogize to anatomical regions, and the output layer analogizes to a behavioral observable. The attractor network paradigm is for recurrent networks, and activity does not converge to a single neuron and extinguish. Instead, the large network is modeled as an embedding of dynamics with a much lower intrinsic dimension. The simple feedforward classifier suffices for illustration purposes. If the network classifies images of cats and dogs, then all correctly classified images of cats creates activity that flows towards the cat neuron. In this example the initial neural encoding of the input image can be regarded as an initial condition and each attractor neuron has a basin of attraction (set of initial conditions) that all flow towards the neuron. Thus, the parameters of the system are fixed and only the initial conditions vary. However, consider an observation of a set of intermediate layers only. Initial conditions are not included in the observation. Nonetheless, activity will still flow towards one particular neuron. By creating a predictive model specific to one observation an observer can guess which output neuron will be activated and guess whether the initial condition corresponded to a cat or a dog. Even in cases where knowledge about the output layer is missing, or the network does not have a simple readout mechanism, an appropriate predictive model would nonetheless separate activity according to initial conditions because initial conditions must produce distinct trajectories. A classifier network with indistinct trajectories in intermediate layers cannot serve its purpose.

Dynamical systems and attractor network interpretations of neural code has seen successful application in population recordings for decades (Skarda and Freeman, 1987; Ben-Yishai et al., 1995; Seung, 1996; Laurent et al., 2001; Laurent, 2002; Wills et al., 2005; Wagenaar et al., 2006; Daelli and Treves, 2010; Gallego et al., 2017, 2020) and new deep learning methods provide a means to identify initial conditions (Pandarinath et al., 2017). One of the key findings has been that the full enumeration of the activity of a large group of neurons is redundant, the actual neural trajectories can be described by accounting for far fewer variables. This idea, that population activity can be described with fewer variables is known as dimensionality reduction. Dimensionality reduction is common in neuroscience, it's implemented through a change of basis or nonlinear transformation followed by a change of basis. In every case the goal is to gain a representation of neural activity in which trajectories have distinct features evincing external variables (e.g. an initial condition like an image presentation). In the simplest case, or after ideal dimensionality reduction, these features would simply be the regions they occupy, a concept called separability. Trajectories corresponding to cats would occupy completely separate regions from trajectories corresponding to dogs. However, this may be too simplistic. For example, in the motor cortex a perfect and complete interpretation of motor neuron activity would produce trajectories that correspond to a motion such as a reaching task because reaching is itself a trajectory (Georgopoulos et al., 1986). Since different reaching motions can intersect, perfectly decoded neural trajectories should be free to intersect and are therefore not separable. Furthermore, cortical networks are highly recurrent meaning that an external stimulus sets up continuing activity with some degree of reverberation. Thus, it can be difficult to unambiguously define an initial condition (Lainscsek et al., 2015). A more

general concept is the context of a recording of neural activity. Context may be defined as any external observable such as stimulus or behavior. Because separability and initial conditions are not broadly generalizable, alternative interpretations of attractor-based encoding are needed. The need for alternative attractor network methods motivated the development of dynamical discrimination. As explained in chapter three, even if the true parameters of a dynamical system are fixed, approximations will show variability if they are fitted to behavior in a small region of state-space or a brief portion of a single trajectory. This is the property we exploit to define a basis for examining the neural code of attractor networks.

Scale-Freeness Implies that Large Scale Neural Dynamics may be Efficiently Represented in Small-Scale Activity

Having recounted the attractor network description for unfolding neural activity and the neuroscientific realities complicating identification of attractor network based neural code we can communicate the significance and impact of current and future work with dynamical discrimination. The study of criticality and dynamical systems have an important point of intersection. For many nonlinear dynamical systems changes to their parameters can cause large changes in their behavior. In particular, a dynamical system may contain points in state space where the derivative vanishes, these are called fixed points. Behavior near a fixed point can converge towards it (in which case it is an attractor), diverge away from in it in all directions, or diverge away in just a few directions (called a saddle point) or can oscillate around it. If one or more parameters are changed and then the number of fixed points change, or the behavior of a fixed-point changes then

those parameter value changes produce a “bifurcation” (Strogatz, 2018). The exact values ODE parameters at the point of change defines the critical point of a bifurcation. This analogizes to phase transitions. Scale-freeness also emerges at the critical point of a dynamical system bifurcation and RG procedures are defined for both bifurcations and phase transitions (Sfondrini, 2012; Gurau et al., 2014).

The attractor network paradigm is the primary application of dynamical systems to neural code. Because the origins of criticality are still undetermined possessing the ability to study both criticality and attractor networks with the same dataset is unprecedented. With the advent of dynamical discrimination this is now possible in whole-cell recordings as well as population recordings. Reasons to combine attractor network theory with the study of criticality go beyond the superficial. As explained above and in chapter two, an RG operator, which analogizes to neural-action, is a basic kind of classifier. Recall that RG operators coarse grain and rescale systems and for systems at the critical point they make statistically identical copies. For the kinds of systems studied with RG operators physicists define an “order parameter” that identifies which side of a phase transition the system is on. If a system is not at the critical point of a phase transition but is extremely close it will nonetheless be to one side or the other of the phase transition. Iterative applications of an RG operator produce systems that are successively more clearly away from the phase transition as indicated by the order parameter. Thus, after enough iterations the order parameter can act as a label indicating the phase of the original system. If a layer of a neural network is RG-like and recurrently connected, then as time passes activity becomes successively more distinct and characteristic of the initial conditions or context.

The most motivating element of this comparison between the study of neural criticality and attractor networks is the question of what happens if a neural network is provided uncertain inputs (e.g. an image of an octopus in a network only trained for cats and dogs). The RG paradigm suggests this would be like iterating on a critical system, each time point of neural activity would be just as inscrutable as the first but would never repeat. The attractor network paradigm would suggest that the activity would never converge to an attractor (since there are only cat attractors and dog attractors in this analogy). In more complicated systems an attractor network may assign a default attractor. If no default attractor is defined, then the only kind of meandering never repeating activity known to dynamical systems is chaos, which is also associated with the critical point of a bifurcation. This association between ambiguous inputs or context and signatures of criticality has evidence to support it. Signatures of criticality may not be universally present. Researchers have repeatedly found that these signatures are stronger and easier to detect during spontaneous activity (Hahn et al., 2010, 2017; Arviv et al., 2015). Because of the close parallels between the properties of RG operations and neural action, and the close parallels between phase transitions and bifurcations it seems that criticality and the neural code intersect in studies of a single-neuron's perspective of the neural code. The work presented in this dissertation may be incremental in unlocking the mysteries of the neural code.

4.1.3 Aim Three: Demonstrate the Appropriate Use of Machine-Learning to Test Scientific Hypotheses

Aim three is separate from the scientific questions about criticality and neural code. Our third aim is important to the practice of science in general but methods for finding neural

code more specifically. Most studies of neural code involve machine learning to perform the actual decoding that links a representation of neural activity to the external variable presumed to be encoded in that activity. There are a handful of machine learning algorithms used for this purpose, few are intended as models of brain function. Consider the popular support vector machine (SVM) algorithm (Cristianini et al., 2000). The SVM defines a function relating each point of data inputted to it to a hyperplane. It then varies the choice of hyperplane to optimize a function. In the case of binary classification, the function is optimum when the hyperplane efficiently separates the two classes. Data points that are above the plane are predicted to be of one class, and data points below it are predicted to be the other. If the classes are not able to be separated with a flat plane the SVM can be extended to include non-linear transformations of the original data to represent the data in a way that is linearly separable. The SVM has seen great success when applied directly to neural data (Mourão-Miranda et al., 2007). The ability of the SVM to perform a quality classification or regression when you provide it one data set, and not another data set indicates that the second data set may lack useful information. Unfortunately, there are many choices in the SVM algorithm, these choices can reverse the situation. What's more, a different algorithm may perform differently all together, even making excellent predictions in both cases. This is a consequence of the famous "no free lunch theorem" (Wolpert, 1996). To completely settle whether two data sets differ in their information content one would have to test all possible forms of all possible algorithms. Even if that were accomplished a core problem remains. Information cannot be created or destroyed, but it can be converted to more or less usable forms. Just because a very clever algorithm proved that neural activity in a specific neural system has information about

some experimental variable does not prove that the information is useful for that system, or even for the researcher. For example, a clever enough algorithm might be able to use the vibrations on the driver-side headrests of cars to precisely predict speed. Such an algorithm tells us nothing about the governing principles of automobiles, even if we compare driver-side headrests to passenger-side headrests. When researchers use the incredible power of modern machine learning, questions about whether one data source can be used to predict another become less useful and potentially misleading. Clearly an alternative way to include machine learning in the process of scientific inference is needed. We provide an example of this approach, which we call hypothesis-dependent machine learning.

Hypothesis-dependent machine learning prescribes that researchers create bespoke algorithms built around a scientific hypothesis such that the success or failure of the algorithm depends on the applicability of the hypothesis. This is a microcosm of scientific inference itself. In the general practice of science, a hypothesis is proposed then predictions are made and then they are tested. If the predictions are wrong, then the hypothesis is revised and tested again. This repeats until a test either provides a fundamental flaw with the hypothesis, or until the hypothesis evolves into a theory that works. A hypothesis dependent machine learning algorithm is limited to tweaking the parameters of a model (i.e. a hypothesis) describing raw data and then using the model for a quantifiable purpose and using the results to inform the next round of tweaks. This is vastly constrained when compared to a non-linear SVM or deep-learning approach which is free to tweak millions, or billions of parameters and find any solution to the problem a researcher codes for it. In

our case the hypothesis was that neural code is describable with dynamical systems theory and the information is contained in the *forces governing* the evolution of neural activity and not necessarily in a snapshot of activity. We compared it to two alternative formulations which depend on snapshots of neural activity, deflection from baseline and a silhouette method: Maximum Likelihood Estimation given the shape of an inferred neural activity trajectory. The complexity of the machine learning algorithm in its entirety was separated from the simplistic core element: modeling neural trajectories with ODEs. Thus, we were able to test principles of neural code by using carefully crafted and useful representations of neural data. These representations (coefficients of ODEs) are certainly not useful to the neurons themselves (we showed that the neuron's presumed function as a feature detector is consistent with deflection from baseline). Regardless, the additional benefits of hypothesis-dependent machine learning, which we invite explore, stem from its utility in revising the hypothesis as the human defines it but with super-human flexibility. Our ODE coefficient matrices showed striking similarities across cells which we have only begun to investigate and interpret. Therefore, our dynamical discrimination algorithm provides an example of leveraging machine learning to meet the unique challenges of data analysis in neuroscience while avoiding the pitfalls of over-interpretation.

4.2 Future directions

The work presented in this text prepares fertile ground for future investigations. Our dynamical discrimination algorithm is ready to be applied to neural data in experiments that may illuminate the effects of criticality and sharpen the focus on neural code. There is

room for improvement in dynamical discrimination itself that promises to enable deeper and more varied insights into dynamical systems perspectives on brain function. While critical averaging, an insight emerging from our modeling work, parallels many important themes in both the study of neural code and criticality and is ready for additional investigations.

4.2.1 Expanding Neural Code and Criticality Comparisons

We have established that whole-cell recordings can evince both subtle changes in the factors governing neural trajectories and indistinct signatures of cortical state. The next course of action is to apply the same methods to more situations. Since the ODE models we fitted are amenable to linear stability analysis (Strogatz, 2018), (the main vehicle by which bifurcations are detected in analysis of nonlinear dynamics) then it makes sense to combine them with studies of criticality. For population recordings where criticality is believed to be modulated by an experimental manipulation the side-by-side comparison of fitted ODEs to measures of consistency with criticality would be both straightforward and informative. Earlier we highlighted similarities between the critical point of bifurcations and phase transitions more broadly. In chapter three we showed that the coefficients of ODEs may reveal distinct dynamical regimes because the coefficients of ODEs fitted to trajectories coinciding the with least and most preferred stimulus led to greater discriminability than for trajectories paired with less salient stimuli. Thus, ODE coefficients that allow the prediction of the presence of criticality would indicate that criticality in

neural systems has measurable consequences for the attractor network paradigm and may even indicate that criticality is dynamical in nature.

More generally the prospect of applying dynamical discrimination to dimensionality reduced population recordings affords a new method for testing the attractor network paradigm in situations where it hasn't been applied before. All applications of dynamical discrimination involve the choice of hyperparameters such as the number of dimensions to include. This is a time-consuming process, but we show how to do it in chapter three. One of these situations is simultaneous or dual recordings. The ODE's used for dynamical discrimination can accept driving terms. Therefore, a population recording in one region paired with an LFP recording in a distant region can be used to understand the impact of the distant region on the factors governing behavior in the region subject to population recording. To do this a research would fit an ODE to intracellular recordings but the ODE would have an extra set of variables corresponding to the LFP recording. This is similar to how injected current is an additional variable in the two-dimensional Fitzhugh-Nagumo ODE.

Applying dynamical discrimination to population recordings also offers the ability to test neural code theories if paired with whole-cell recordings (not necessarily simultaneously). If an activity pattern of any kind is observed in a population and believed to be a useful component of the brain's intrinsic neural code, then dynamical discrimination (or other approach) should be able to predict the occurrences of that pattern from intracellular recordings of neurons that receive most of their inputs from that population. Often these

neurons are within that same population, but not necessarily. The format of these predictions could be simple labels (e.g. epochs where one community or ensemble dominates). Alternatively, a researcher could fit ODEs to population recordings and map them to ODEs fitted to intracellular recordings. If done asynchronously (or in different animals) researcher would then repeat the experiment with and without the conditions designed to elicit the pattern they thought they had observed and use dynamical discrimination on whole-cell recordings to predict whether those conditions were present. If the relationship between the ability to discriminate conditions matches the presence or absence of the pattern, then it suggests that the pattern is strongly represented in intrinsic neural representations. If done synchronously a researcher can directly predict the pattern from the ODE coefficients. One additional, promising approach is to create a time-series variable that summarizes the population activity. This variable could be integers that denote which ensemble is present, or a multivariate time-series of the average activity of more than one ensemble. The researcher would then use it as an input when fitting ODE coefficients (just like the LFP case) and explore how and whether it changes the form and accuracy of the fitted ODEs.

4.2.2 Improving and Extending Dynamical Discrimination Methods

There are many open questions and areas for improvement with dynamical discrimination. Chief among these would be the ability to interpret the ODE equations. We intentionally fit to brief snippets to get crude approximations because this led to better variability and consequently, discrimination. Thus, there is little expectation that these are universal models of neural dynamics. Nonetheless, the trajectories were oscillatory and, when

viewed straight down the axis of rotation, looked somewhat like Fitzhugh-Nagumo phase portraits (Izhikevich and FitzHugh, 2006). We tried to drive Fitzhugh-Nagumo model neurons with known sinusoidal dynamics in the hopes that dimensionality expansion would clearly separate them, with two dimensions corresponding to the Fitzhugh-Nagumo variables and two dimensions corresponding to the sinusoid variables. We found that the dimensions mixed and could not be easily separated. We tried some attempts to learn an orthogonal projection matrix as an alternative to PCA that let us project the time-delay over-expansion (see chapter three) onto four dimensions. The goal was to have Fitzhugh-Nagumo equations fit well to the first two dimensions and leave the remaining two free while still concentrating the variance. We hoped that this would lets us fit ODEs that captured the Fitzhugh-Nagumo and the sinusoid coefficients. These attempts were unsuccessful, but if methods could be found to carry out such a transformation then it would be the first step to intentionally separating intrinsic neural dynamics from input dynamics. The ODEs governing input dynamics would then be suspected to match ODEs that might be fitted to the input population. Since we never fitted ODEs to an input population this could still be the case.

Another approach to improving the interpretability of the ODEs is a small change to the genetic algorithm that fits them. There are four ways to do this. Currently we perform classification by looking at all ODE coefficients. However, human beings favor analysis of dynamical models in which only one or two parameters are related to an experimental variable. An example would be a coefficient whose magnitude changes as the sine of the orientation of a drifting grating. The first way to produce this effect would be to force the

genetic algorithm to make predictions using just one or two coefficients. Currently we predict drifting gratings from the entire set of coefficients. An experimenter might then plot a relationship between the coefficients that the genetic algorithm has isolated and the stimulus features that varied. A second way to improve the interpretability of the ODEs would be for an experimenter to make a guess about an ODE form and force the genetic algorithm to include it in its fits. A third way is to measure the mutual information between each coefficient and the stimulus variable. We did his but did not find any coefficients that stood out. If we had it would, we would have tried to define a relationship between the value of the coefficient and drifting grating variables. Lastly, the fourth way to improve the interpretability of the ODEs would be to change how classifications are made. Rather than learning one ODE for all recordings for each cell (pooling all the trials) the genetic algorithm might learn one ODE for each group of recordings coinciding with each stimulus (separating trials by stimulus). After doing this a researcher might use the time-points of an unexamined recording (a holdout set) as initial conditions for integrating each of the stimulus-matched ODEs. The ODE that makes the best prediction corresponds to the most likely stimulus (or other external variable). We tried this using the coefficient of determination between real and predicted derivatives but did not get satisfactory results. However, it was early in algorithm development and we made other improvements that may see this approach work. Additionally, one could use different standards. One could modify the measure of goodness of fit to see how well the ODE predicts the actual trajectory, rather than trajectory derivatives. One could, see what fraction of initial conditions diverged or converged to extinguishing points. If an algorithm based on this

approach succeeds then the analyst will have uncovered different ODEs for different conditions and may be able to discover a relationship between them.

4.2.3 Critical Averaging Robustness and Relationship to Homeostasis and General Computation

A key aspect of our modeling investigation provided an understanding for why the profiles of single cell membrane potential fluctuations would closely match the profiles of neuronal avalanches. We demonstrated that the same condition that gives rise to critical branching in our model also produced what we called "critical averaging" where each neuron's instantaneous firing rate approximates the population firing rate of its inputs. Importantly we found that the critical averaging condition held even when we could not show this for the critical branching condition (e.g. in the presence of inhibition). Even though this condition is the same as the condition for criticality in models without inhibition it is not analytically demonstrated that it is sufficient to produce criticality in models with inhibition or with different connectivity structure. Numerical modeling showed that networks with inhibition are likely to be critical. Furthermore, critical averaging is by definition a self-regulating constraint that parallels models of homeostasis (Hsu and Beggs, 2006; Pozo and Goda, 2010), and it is also a computational relationship between the inputs to a neuron and its output. Therefore, important future work would be to explore four important topics. First, whether the critical averaging condition is robust to diverse network architectures such as highly clustered, small world and the distribution of inhibitory neurons. Second, whether the critical averaging condition analytically implies neuronal avalanches that obey the power-law predictions of criticality. Third, whether

critical averaging can be a general result of plasticity mechanisms especially those associated with homeostasis and criticality. Fourth, analytical and modeling studies into whether a network whose neurons are constrained to a critical averaging type neural-action can serve as universal function approximators and whether these networks have universal properties consistent with RG and critical systems theory.

The first and second additional topics, the robustness of the critical averaging condition to diverse network architectures and whether it analytically implies signatures of criticality, can be investigated quite efficiently. In chapter two we use elementary mathematical statistics to capture the expected values and variability of piecewise uniform distributions. These distributions govern connection probability and connection strength. To ascertain whether the firing rate of a single neuron approximates its inputs we calculated the expected firing probability given these distributions and patterns of spiking. This mirrors work that originally established the dual condition (critical branching) and demonstrated it yields a directed percolation critical system (Larremore et al., 2011, 2012). Different network architectures correspond to different connection probability distributions (Newman, 2018). Hence testing network architectures can proceed similarly with elementary mathematical statistics (Gubner, 2006). The first step would be to show there exists a set of parameter choices that allow critical averaging, this is sufficient to establish robustness. The next (and last) step is either to show that these parameter choices results in avalanches that have the power law size and duration distributions associated with directed percolation, or to show that the critical branching condition is also satisfied. In order to demonstrate critical branching in a clustered network one would have to consider

that neurons in the same cluster are more correlated relative neurons in different clusters. Having analytical results for these claims would greatly bolster the promising model work we have initially conducted.

The third and fourth topics, connecting critical averaging to plasticity mechanisms and computational paradigms permit analytical solutions, but modeling is an ideal strategy for investigation. Many plasticity mechanisms such as STDP obey differential equations that approach an equilibrium and may permit quasiperiodic solutions (Effenberger et al., 2015; Soloduchin and Shamir, 2018). Because they use more complicated neuron functions (LIF, Izhikevich, etc.) finding the firing rate relationships implied by the steady state weights involves solving high-dimensional dynamical systems that likely do not permit unique solutions. Therefore, a modeling approach is prudent. Once a researcher has found plasticity mechanisms that produce critical averaging, they would then perturb the network and investigate homeostasis. It would be important to test whether critical averaging is consistent with multiple homeostatic mechanisms. A popular mechanism is synaptic scaling (Turrigiano, 2008), which proposes that synaptic strengths are modulated such that the postsynaptic neuron's firing rate neither greatly exceeds nor falls short of the presynaptic population firing rate. This seems likely to produce critical averaging but is not the only way. For critical average to produce conditions congruent with critical branching, a neuron's firing rate must only approximate the firing rate of *any* other group. It is reasonable to assume that the group forms direct inputs to that neuron, but this isn't necessarily the case. Because the matter has not been investigated there is no reason to discount the possibility that the group governing the firing rate of a given neuron could

vary in time. Thus, many homeostatic mechanisms, even those that are not cell-intrinsic (Marder and Prinz, 2002; Turrigiano, 2011; Benucci et al., 2013; Hengen et al., 2013) may nonetheless result in critical averaging. Should critical averaging prove a useful concept for neural computation it would also be advantageous to understand how critical averaging could work with groups far removed from the neuron of interest. We just outlined ways to investigate the third critical-averaging related topic, next we explore how to investigate the computational implications.

Two computational ideas may play a role in understanding how a seemingly simple effect, like critical averaging, could result from rich and complex neural-actions. The first is the venerable idea of neurons as feature detectors in the visual system (Martin, 1994). Should a group of neurons be more likely to fire when the feature is more likely to be present then their individual firing rates may be unreliable, but their average firing rate becomes more reliable as the size of the group increases. For more complicated visual features that may be composites of rudimentary features an average across multiple simple-feature-associated neural groups reveals the likelihood of the composite feature. The idea of feature detectors has been challenged and elaborated on in the decades since it was first proposed resulting in computational paradigms such as predictive coding or efficient coding (Friston and Kiebel, 2009; Huang and Rao, 2011; Denève and Machens, 2016) but it serves as a basis for understanding how an averaging operation can perform useful tasks such as distilling information about a visual scene.

The second computational idea that may help understand how a simple effect like critical averaging can fill a complex computational need is borrowed from the study of large-scale brain networks. The searchlight hypothesis (Crick, 1984; McAlonan et al., 2008) which proposes that feature integration does not operate on the entire visual scene at once, but rather the visual system directs its energy to important regions. There is not a priori reason the concept of time-varying selectivity can't be applied to smaller scale systems. Neuroscience has already discovered that the sensitivity of individual branches of the dendritic arborization can be modulated (Branco and Häusser, 2010; Legenstein and Maass, 2011). Thus, complex phenomena within the dendritic arborization can shift the averaging process between various subgroups. In other words, a neuron's firing probability might match that of one sub-group of its inputs in one moment and then match the firing probability of a different sub-group in a different moment. Should the critical averaging condition be robust to this kind of time-varying aggregation then it would be plausible as both a general computational concept and an origin for the observations of criticality. To test whether critical averaging is compatible with general computation a researcher would implement it in neural networks designed to perform computations. Since critical average has a simple implementation in the models used in chapter one (weights rescaling) it is elementary to both enforce this condition and define backpropagation for both feedforward and recurrent versions. One simply switches from stochastic firing events to a rectified linear neuron activation function which makes our model very similar to the Restricted Boltzmann Machine (Nair and Hinton, 2010). Thus, it can be trained to perform computations. Then the researcher switches back to stochastic neural activation. Existing methods can also be adapted to test time varying aggregation. Current recurrent neural

networks exist which can detect sudden changes in their inputs. Researchers can use that feature to switch between different classification networks that all have critical averaging implemented. Regardless of classification, the idea of switching network connectivity structures in the midst of an ongoing simulation is one way to test the robustness of critical averaging to time-varying aggregation. Examining the spiking statistics of the combined networks for consistency with criticality would be an effective test of the critical averaging condition.

4.3 Conclusion

In conclusion, we have satisfied our three aims: To expand and clarify the study of scale-freeness and criticality in neural systems, to articulate intrinsic representations of neural dynamics and neural code, and to demonstrate the appropriate use of machine-learning to test scientific hypotheses. The work in this dissertation, and the software to reproduce it lays a solid foundation for progress. Critical averaging offers a coarse-graining based perspective on computation that may allow fruitful insights into the origins of scale-freeness, homeostasis, and self-organized criticality. Dynamical discrimination offers a simple tool for investigating the attractor network paradigm for neural code. With these works, and these prospects physics and neuroscience have come together and will continue to come together as we investigate criticality in neural systems and dynamical attractor based neural codes.

References

Amit DJ, Amit DJ (1992) Modeling Brain Function: The World of Attractor Neural Networks. Cambridge University Press. Available at: https://play.google.com/store/books/details?id=fvLYch1yQncC.

Arviv O, Goldstein A, Shriki O (2015) Near-Critical Dynamics in Stimulus-Evoked Activity of the Human Brain and Its Relation to Spontaneous Resting-State Activity. J Neurosci 35:13927–13942 Available at: http://dx.doi.org/10.1523/JNEUROSCI.0477-15.2015.

Beggs JM, Plenz D (2003) Neuronal avalanches in neocortical circuits. J Neurosci 23:11167–11177 Available at: https://www.ncbi.nlm.nih.gov/pubmed/14657176.

Benucci A, Saleem AB, Carandini M (2013) Adaptation maintains population homeostasis in primary visual cortex. Nat Neurosci 16:724–729 Available at: http://dx.doi.org/10.1038/nn.3382.

Ben-Yishai R, Bar-Or RL, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. Proc Natl Acad Sci U S A 92:3844–3848 Available at: http://dx.doi.org/10.1073/pnas.92.9.3844.

Branco T, Häusser M (2010) The single dendritic branch as a fundamental functional unit in the nervous system. Curr Opin Neurobiol 20:494–502 Available at: http://dx.doi.org/10.1016/j.conb.2010.07.009.

Brunel N, Hakim V, Richardson MJE (2014) Single neuron dynamics and computation. Curr Opin Neurobiol 25:149–155 Available at: http://dx.doi.org/10.1016/j.conb.2014.01.005.

Brunton SL, Proctor JL, Kutz JN (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc Natl Acad Sci U S A 113:3932–3937 Available at: http://dx.doi.org/10.1073/pnas.1517384113.

Buice MA, Cowan JD (2009) Statistical mechanics of the neocortex. Prog Biophys Mol Biol 99:53–86 Available at: http://dx.doi.org/10.1016/j.pbiomolbio.2009.07.003.

Crick F (1984) Function of the thalamic reticular complex: the searchlight hypothesis. Proc Natl Acad Sci U S A 81:4586–4590 Available at: http://dx.doi.org/10.1073/pnas.81.14.4586.

Cristianini N, Shawe-Taylor J, Department of Computer Science Royal Holloway John Shawe-Taylor (2000) An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press. Available at: https://play.google.com/store/books/details?id=_PXJn_cxv0AC.

Daelli V, Treves A (2010) Neural attractor dynamics in object recognition. Exp Brain Res 203:241–248 Available at: http://dx.doi.org/10.1007/s00221-010-2243-1.

Denève S, Machens CK (2016) Efficient codes and balanced networks. Nat Neurosci 19:375–382 Available at: http://dx.doi.org/10.1038/nn.4243.

Effenberger F, Jost J, Levina A (2015) Self-organization in Balanced State Networks by STDP and Homeostatic Plasticity. PLoS Comput Biol 11:e1004420 Available at: http://dx.doi.org/10.1371/journal.pcbi.1004420.

Eliasmith C (2007) Attractor network. Scholarpedia J 2:1380 Available at: http://www.scholarpedia.org/article/Attractor_network.

Friston K, Kiebel S (2009) Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci 364:1211–1221 Available at: http://dx.doi.org/10.1098/rstb.2008.0300.

Gallego JA, Perich MG, Chowdhury RH, Solla SA, Miller LE (2020) Long-term stability of cortical population dynamics underlying consistent behavior. Nat Neurosci Available at: http://dx.doi.org/10.1038/s41593-019-0555-4.

Gallego JA, Perich MG, Miller LE, Solla SA (2017) Neural Manifolds for the Control of Movement. Neuron 94:978–984 Available at: http://dx.doi.org/10.1016/j.neuron.2017.05.025.

Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. Science 233:1416–1419 Available at: http://dx.doi.org/10.1126/science.3749885.

Gidon A, Zolnik TA, Fidzinski P, Bolduan F, Papoutsi A, Poirazi P, Holtkamp M, Vida I, Larkum ME (2020) Dendritic action potentials and computation in human layer 2/3 cortical neurons. Science 367:83–87 Available at: http://dx.doi.org/10.1126/science.aax6239.

Gubner JA (2006) Probability and Random Processes for Electrical and Computer Engineers. Cambridge University Press. Available at: https://play.google.com/store/books/details?id=pa20eZJe4LIC.

Gurau R, Rivasseau V, Sfondrini A (2014) Renormalization: an advanced overview. arXiv [hep-th] Available at: http://arxiv.org/abs/1401.5003.

Hahn G, Petermann T, Havenith MN, Yu S, Singer W, Plenz D, Nikolic D (2010) Neuronal avalanches in spontaneous activity in vivo. J Neurophysiol 104:3312–3322 Available at: http://dx.doi.org/10.1152/jn.00953.2009.

Hahn G, Ponce-Alvarez A, Monier C, Benvenuti G, Kumar A, Chavane F, Deco G, Frégnac Y (2017) Spontaneous cortical activity is transiently poised close to criticality. PLoS

Comput Biol 13:e1005543 Available at: http://dx.doi.org/10.1371/journal.pcbi.1005543.

He BJ (2014) Scale-free brain activity: past, present, and future. Trends Cogn Sci 18:480–487 Available at: http://dx.doi.org/10.1016/j.tics.2014.04.003.

Hengen KB, Lambo ME, Van Hooser SD, Katz DB, Turrigiano GG (2013) Firing rate homeostasis in visual cortex of freely behaving rodents. Neuron 80:335–342 Available at: http://dx.doi.org/10.1016/j.neuron.2013.08.038.

Hertz J (1995) Computing with attractors. The handbook of brain theory and neural networks:230–234 Available at: http://www.nbi.dk/~hertz/papers/att02.ps.gz.

Hsu D, Beggs JM (2006) Neuronal avalanches and criticality: A dynamical model for homeostasis. Neurocomputing 69:1134–1136 Available at: http://www.sciencedirect.com/science/article/pii/S0925231205003863.

Huang Y, Rao RPN (2011) Predictive coding. Wiley Interdiscip Rev Cogn Sci 2:580–593 Available at: http://dx.doi.org/10.1002/wcs.142.

Izhikevich EM, FitzHugh R (2006) Fitzhugh-nagumo model. Scholarpedia J 1:1349 Available at: http://www.scholarpedia.org/article/Bonhoeffer-van_der_Pol.

Koch-Janusz M, Ringel Z (2018) Mutual information, neural networks and the renormalization group. Nat Phys 14:578–582 Available at: https://doi.org/10.1038/s41567-018-0081-4.

Lainscsek C, Weyhenmeyer J, Sejnowski TJ, Letellier C (2015) Discovering independent parameters in complex dynamical systems. Chaos Solitons Fractals 76:182–189 Available at: http://dx.doi.org/10.1016/j.chaos.2015.04.001.

Larremore DB, Carpenter MY, Ott E, Restrepo JG (2012) Statistical properties of avalanches in networks. Phys Rev E Stat Nonlin Soft Matter Phys 85:066131 Available at: http://dx.doi.org/10.1103/PhysRevE.85.066131.

Larremore DB, Shew WL, Ott E, Restrepo JG (2011) Effects of network topology, transmission delays, and refractoriness on the response of coupled excitable systems to a stochastic stimulus. Chaos 21:025117 Available at: http://dx.doi.org/10.1063/1.3600760.

Laurent G (2002) Olfactory network dynamics and the coding of multidimensional signals. Nat Rev Neurosci 3:884–895 Available at: http://dx.doi.org/10.1038/nrn964.

Laurent G, Stopfer M, Friedrich RW, Rabinovich MI, Volkovskii A, Abarbanel HD (2001) Odor encoding as an active, dynamical process: experiments, computation, and theory. Annu Rev Neurosci 24:263–297 Available at: http://dx.doi.org/10.1146/annurev.neuro.24.1.263.

Legenstein R, Maass W (2011) Branch-specific plasticity enables self-organization of nonlinear computation in single neurons. J Neurosci 31:10787–10802 Available at: http://dx.doi.org/10.1523/JNEUROSCI.5684-10.2011.

Li S-H, Wang L (2018) Neural Network Renormalization Group. Phys Rev Lett 121:260601 Available at: http://dx.doi.org/10.1103/PhysRevLett.121.260601.

London M, Häusser M (2005) Dendritic computation. Annu Rev Neurosci 28:503–532 Available at: http://dx.doi.org/10.1146/annurev.neuro.28.061604.135703.

Marder E, Prinz AA (2002) Modeling stability in neuron and network function: the role of activity in homeostasis. Bioessays 24:1145–1154 Available at: http://dx.doi.org/10.1002/bies.10185.

Martin KAC (1994) A Brief History of the "Feature Detector." Cereb Cortex 4:1–7 Available at: https://academic.oup.com/cercor/article-abstract/4/1/1/440519 [Accessed January 24, 2020].

Ma Z, Turrigiano GG, Wessel R, Hengen KB (2019) Cortical Circuit Dynamics Are Homeostatically Tuned to Criticality In Vivo. Neuron 104:655–664.e4 Available at: http://dx.doi.org/10.1016/j.neuron.2019.08.031.

McAlonan K, Cavanaugh J, Wurtz RH (2008) Guarding the gateway to cortex with attention in visual thalamus. Nature 456:391–394 Available at: http://dx.doi.org/10.1038/nature07382.

Mehta P, Schwab DJ (2014) An exact mapping between the Variational Renormalization Group and Deep Learning. arXiv [statML] Available at: http://arxiv.org/abs/1410.3831.

Mourão-Miranda J, Friston KJ, Brammer M (2007) Dynamic discrimination analysis: a spatial-temporal SVM. Neuroimage 36:88–99 Available at: http://dx.doi.org/10.1016/j.neuroimage.2007.02.020.

Nair V, Hinton GE (2010) Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 807–814. cs.toronto.edu. Available at: https://www.cs.toronto.edu/~hinton/absps/reluICML.pdf.

Newman M (2018) Networks. Oxford University Press. Available at: https://play.google.com/store/books/details?id=YdZjDwAAQBAJ.

Pandarinath C, Collins J, Jozefowicz R, Stavisky S, Kao J, Churchland M, Kaufman M, Ryu S, Henderson J, Shenoy K, Others (2017) Precise Estimates of Single-Trial Dynamics in Motor Cortex using Deep Learning Techniques. Available at: https://research.google/pubs/pub46132/.

Pozo K, Goda Y (2010) Unraveling mechanisms of homeostatic synaptic plasticity. Neuron 66:337–351 Available at: http://dx.doi.org/10.1016/j.neuron.2010.04.028.

Sethna JP, Dahmen KA, Myers CR (2001) Crackling noise. Nature 410:242–250 Available at: http://dx.doi.org/10.1038/35065675.

Seung HS (1996) How the brain keeps the eyes still. Proc Natl Acad Sci U S A 93:13339–13344 Available at: http://dx.doi.org/10.1073/pnas.93.23.13339.

Sfondrini A (2012) An introduction to universality and renormalization group techniques. arXiv [hep-th] Available at: http://arxiv.org/abs/1210.2262.

Skarda CA, Freeman WJ (1987) How brains make chaos in order to make sense of the world. Behav Brain Sci 10:161–173 Available at: https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/how-brains-make-chaos-in-order-to-make-sense-of-the-world/3811D1F3B8203EFE80CAF7449D0792DF [Accessed January 13, 2020].

Soloduchin S, Shamir M (2018) Rhythmogenesis evolves as a consequence of long-term plasticity of inhibitory synapses. Sci Rep 8:13050 Available at: http://dx.doi.org/10.1038/s41598-018-31412-7.

Strogatz SH (2018) Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. CRC Press. Available at: https://www.taylorfrancis.com/books/9780429492563.

Turrigiano G (2011) Too many cooks? Intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement. Annu Rev Neurosci 34:89–103 Available at: http://dx.doi.org/10.1146/annurev-neuro-060909-153238.

Turrigiano GG (2008) The self-tuning neuron: synaptic scaling of excitatory synapses. Cell 135:422–435 Available at: http://dx.doi.org/10.1016/j.cell.2008.10.008.

Wagenaar DA, Nadasdy Z, Potter SM (2006) Persistent dynamic attractors in activity patterns of cultured neuronal networks. Phys Rev E Stat Nonlin Soft Matter Phys 73:051907 Available at: http://dx.doi.org/10.1103/PhysRevE.73.051907.

Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J (2005) Attractor dynamics in the hippocampal representation of the local environment. Science 308:873–876 Available at: http://dx.doi.org/10.1126/science.1108905.

Wolpert DH (1996) The Lack of A Priori Distinctions Between Learning Algorithms. Neural Comput 8:1341–1390 Available at: http://dx.doi.org/10.1162/neco.1996.8.7.1341.

Appendix

A.1 Dynamical Discrimination as Timeseries Classifier: Performance and Additional Insights

We offer a new method for analyzing electrophysiological data that brings neural population analysis to single whole-cell recordings. However, the same methods can be applied to any single-unit recording and the method can be easily adapted to multi-unit recordings. Applying discrete labels to time series of any dimensionality is a form of trajectory classification (Maharaj et al., 2019). Examples include predicting the final destination of a vehicle based on a small sample of it's path (Lee et al., 2008), or identifying a fundamental particle from its path in a bubble chamber (Hough, 1959). There are as many methods as applications. For neural data the goal is often to predict what stimulus (Laurent, 2002; Mazor and Laurent, 2005; Saha et al., 2013) or behavior (Churchland et al., 2012; Gallego et al., 2017; Pandarinath et al., 2018) co-occurred with the neural recording, but it could also be used to examine the impact of other experimental manipulations or observations. In brief all methods of trajectory classification seek to reduce trajectories to small sets of numbers which either correspond to a physical property (e.g. mass) or which is still abstract but small enough to train a standard classifier on. Since trajectories contain very large amounts of data (one data point for each dimension for each point in time) one would require very many examples if they used the whole trajectories as training data for a classifier such as a support vector machine, random forest, or neural network. These steps

can be avoided by using recurrent neural networks such as LSTM, where it is known as "sequence classification" (Kawakami, 2008). However, this usually requires large amounts of data to train on and is not model-dependent, therefore not easily interpreted or manipulated to gain additional insights. Attempts to use them for our whole cell recordings were not satisfactory. The trajectory classification approach we developed is an exciting new process because an experimentalist can test a hypothesized relationship by testing whether expected information is "encoded" in the dynamical rules governing the trajectory.

The method we use is an adaptation of the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm (Kawakami, 2008; Kaiser et al., 2018). This is rooted in more than convenience but in the attractor network theory of brain function. This framework seeks to model the brain as a network of dynamical nodes (a very high dimensional nonlinear dynamical system). With this perspective any neural response to stimuli with any degree of stereotypy is considered an attractor. The state space actually explored by the attractor is usually found to be a small fraction of the possible state-space, if possible state space is defined as an N-dimensional space where each dimension corresponds to the firing rate of each of the N-neurons. The much smaller subspace which is actually explored is often referred to as a neuronal manifold (Kawakami, 2008; Gallego et al., 2017; Kaiser et al., 2018). Currently the most common way to try to visualize and quantify this low-dimensional manifold is through dimensionality reduction on high-dimensional recordings. The use of nonlinear methods of dimensionality reduction have been able to reduce the dimensionality of neural data much further than linear methods such as PCA or SVD, and arguments have been made that nonlinear dimensionality reduction is a more faithful

representation of the functionally relevant mechanisms (Gallego et al., 2017). An alternative is to expand one-dimensional recordings to a moderate dimensional space consistent with neuronal manifolds. The theory of non-linear dynamics offers some guarantees that this can be done, through the Whitney and Takens delay embedding theorems (Sauer et al., 1991). Delay embedding theorem guarantees that we can capture the high dimensional dynamics by observing any single dimension. Nonetheless it is unclear what information is lost as the neuron samples upstream populations, thus the relationship between dimensionality expansion on whole-cell recordings and the dynamics obtained through dimensionality reduction is unclear. Fortunately our core inference mechanism is estimation of ordinary differential equations (ODEs) and ODEs can be fit to either dimensionally reduced or dimensionally expanded data. Therefore, if an analyst is using dimensionality reduction on multi-unit data and already has trajectories in a putative state space then the methods that follow apply without modification.

In the study of nonlinear dynamics it is usually the case that the dynamics near an attractor are simpler than the dynamics describing the rest of state space. For example, in a system with many dimensions choices of model parameters and/or initial conditions near a limit cycle may be well approximated with simple harmonic motion whereas parameters at a bifurcation, or initial conditions far from a limit cycle may exhibit complex orbits (Strogatz, 2018). This last simplification is what we depend on most for our method to work. We estimate the derivative of each dimension of our trajectories and we test numerous forms of 3rd order polynomial differential equations until we find a system of equations for each cell that has few terms but gives us the best ability to predict what stimulus co-occurred

with each recording of a trial. We also explore models that are focused purely on goodness of fit and not classification. The fitting procedure is a least squares regression between selected polynomial combinations of our trajectory dimensions and the estimated derivatives. However the selection of terms is carried out with a genetic algorithm (see methods). The result is that by taking the attractor computation perspective and fitting crude ODEs to short-duration dynamics we are able to produce a representation (the coefficients of ODEs) of single whole cell recordings compact enough with 12 ± 1.5 real numbers (median $\pm$ half interquartile range) that we can train a classifier with 68 ± 18.75 total trials per cell and only 11 ± 2.5 samples of each type of stimulus for each cell with either 6 or 8 unique stimuli per cell. This representation is small enough that we can now perform classification despite having few trials.

Because we are performing trajectory classification to identify stimulus by finding systems of differential equations we call our process "dynamical discrimination". However, in order to perform the final classification step we need a classifier. Sometimes, (e.g. particle detectors) the compact representation are basic physical properties like mass and charge (Hough, 1959; Strogatz, 2018), and we need only to look it up in a table. In our case we have an abstract compact representation and have to create that look up table ourselves by training a random forest classification algorithm (Hough, 1959; Breiman, 2001; Strogatz, 2018). A random forest is an ensemble of decision trees. Each tree is a straight-forward conditional look-up table. The trees in the forest "vote" on the correct label to apply. This is among the simplest types of classifiers to use and is able to handle nonlinear problems. The trees are trained by taking a subset of the data that has already been labeled and finding

patterns between the labels and the compact representations. The random forest is tested by seeing if those patterns allow a flow chart of true/false tests that ends with the assignment of correct labels for data that was not used for training.

Traditionally, neuroscientists have sought scalar summaries of intracellular recordings with the purpose of predicting a single neuron's firing rate under the assumption that this is the most useful goal of single neuron observation (Wolfe et al., 2010). The membrane potential of our neurons fluctuated between -72 mV and -62 mV in the absence of visual stimulation. Within 100 ms of activating a stimulus (an image presented on a screen) the cell membrane depolarizes by an amount which depends on many factors. The primary factor is whether the stimulus is a "preferred stimulus" for a neuron (Butts and Goldman, 2006). A preferred stimulus is one which is more likely than others to induce action potentials from that neuron. So if a neuron happens not to fire an action potential upon presentation of a preferred stimulus the amount of depolarization of membrane potential should still be greater than for less preferred stimuli. It is well established that the amount of depolarization correlates strongly with firing rate and depolarization can be used as a proxy for firing rate when estimating feature selectivity (Carandini and Ferster, 2000; Butts and Goldman, 2006). This interpretation is frequently carried over into voltage-clamp recordings of inhibitory or excitatory transmembrane current (Ferster, 1986). In these cases the term "deflection" rather than depolarization is more general so for simplicity we use deflection when discussing all signals. Deflection captures the difference in mean signal value during and not-during stimulus presentation. By using deflection we

can compare with the previous accomplishments and understanding of whole-cell recording analysis which used the same or similar measures (Adesnik, 2017).

The following table breaks results down the results of our dynamical discrimination algorithm by each kind of recording and stimulus. Given the cell -to-cell variability that exists it is necessary to judge whether the median value is indicative of the category behavior and one can expect to get the same results with other data sets. To that end we subtract the chance level of performance and use the Wilcoxon signed rank test (Wilcoxon, 1992) to judge whether the median is better than chance and report on the effect size r_{sdf} (simple difference formula (Kerby, 2014) and its p-value. This table verifies the annotations of significance reported in figures 3, 4, and A.9.

category	algorithm	median CCR	N	dynamical discrimination (DynDisc)		best-fit Ξ based discrimination (Best-Fit)		deflection based discrimination (DefDisc)		Maximum Likelihood Estimation (MLE)		dynamically stable state discrimination (StateDisc)		chance	
				r_{sdf}	p-value	r_{sdf}	p-value	r_{sdf}	p-value	r_{sdf}	p-value	r_{sdf}	p-value	r_{sdf}	p-value
All	DynDisc	0.2875	110			0.1542	0.01893	0.2022	3.60E-08	0.1948	1.03E-07	0.1409	6.05E-03	0.2385	5.49E-27
	Best-Fit	0.2797	110	0.09696	0.9812			0.1861	1.54E-05	0.1807	2.35E-05	0.1259	0.4945	0.2427	8.92E-24
	DefDisc	0.2281	95	0.04908	1	0.0652	1			0.1182	0.7382	0.06007	1	0.2218	5.85E-14
	MLE	0.2481	110	0.05633	1	0.07046	1	0.1351	0.263			0.06903	1	0.2246	2.04E-15
	StateDisc	0.2833	95	0.07487	0.994	0.1254	0.5063	0.1912	2.73E-06	0.1823	5.04E-05			0.2377	8.28E-24
I,O|I	DynDisc	0.1437	7			0.219	0.07813	0.1905	0.1875	0.2571	0.01563	0.2095	0.1172	0.2429	0.03125
	Best-Fit	0.1256	7	0.04762	0.9453			0.1429	0.4688	0.219	0.07813	0.09524	0.7656	0.1619	0.3438
	DefDisc	0.1353	7	0.07619	0.8516	0.1238	0.5938			0.1905	0.1875	0.1333	0.5313	0.1619	0.1094
	MLE	0.08612	7	9.52E-03	0.9922	0.04762	0.9453	0.07619	0.8516			0.04762	0.9453	0.1333	0.5313
	StateDisc	0.1375	7	0.05714	0.9297	0.1714	0.2891	0.1333	0.5313	0.219	0.07813			0.2143	0.08594
E,O|I	DynDisc	0.2562	7			0.2476	0.02344	0.2571	0.01563	0.2476	0.02344	0.181	0.04688	0.2	0.01563
	Best-Fit	0.1908	7	0.01905	0.9844			0.2381	0.03906	0.1714	0.2891	0.08571	0.8125	0.2381	0.03906
	DefDisc	0.1431	7	9.52E-03	0.9922	0.02857	0.9766			0.0381	0.9609	9.52E-03	0.9922	0.1429	0.4688
	MLE	0.1503	7	0.01905	0.9844	0.09524	0.7656	0.2286	0.05469			0.08571	0.8125	0.2095	0.1094
	StateDisc	0.2062	7	0.01905	0.9688	0.181	0.2344	0.2571	0.01563	0.181	0.2344			0.1905	0.03125
I,C|I	DynDisc	0.3167	15			0.1398	0.402	0.1753	0.1174	0.1828	0.08441	0.05484	0.9186	0.2581	3.05E-05
	Best-Fit	0.3255	15	0.1183	0.6192			0.1763	0.1147	0.1914	0.0535	0.06667	0.9527	0.2581	3.05E-05
	DefDisc	0.3253	15	0.0828	0.8881	0.08172	0.8961			0.1677	0.1651	0.05914	0.9692	0.2581	3.05E-05
	MLE	0.2954	15	0.07527	0.9243	0.06667	0.9527	0.09032	0.8486			0.03656	0.9949	0.2581	3.05E-05
	StateDisc	0.35	15	0.1409	0.08435	0.1914	0.0535	0.1989	0.03296	0.2215	6.23E-03			0.2581	3.05E-05
E,C|I	DynDisc	0.2917	15			0.1226	0.5765	0.1591	0.2271	0.1742	0.1262	0.07312	0.8778	0.2581	3.05E-05
	Best-Fit	0.2917	15	0.1355	0.4452			0.1656	0.1796	0.1763	0.1147	0.09032	0.8486	0.2581	3.05E-05
	DefDisc	0.249	15	0.09892	0.7894	0.09247	0.8349			0.1333	0.467	0.06667	0.9527	0.2559	6.10E-05
	MLE	0.2281	15	0.08387	0.8853	0.08172	0.8961	0.1247	0.5548			0.04946	0.9849	0.2065	0.02063
	StateDisc	0.3167	15	0.1527	0.1301	0.1677	0.1651	0.1914	0.0535	0.2086	0.01787			0.2581	3.05E-05
R,C|V	DynDisc	0.275	10			0.1714	0.2158	0.1571	0.3125	0.1333	0.5	0.1333	0.1016	0.2238	0.02246
	Best-Fit	0.2372	10	0.09048	0.8125			0.1095	0.6875	0.119	0.6152	0.1667	0.2461	0.2476	4.88E-03
	DefDisc	0.2463	10	0.1048	0.7217	0.1524	0.3477			0.1333	0.5	0.1762	0.1875	0.2524	2.93E-03
	MLE	0.2455	10	0.1286	0.5391	0.1429	0.4229	0.1286	0.5391			0.1905	0.1162	0.2476	4.88E-03
	StateDisc	0.2042	10	0.0381	0.9297	0.09524	0.7842	0.08571	0.8389	0.07143	0.9033			0.2143	0.04492
I,S|I	DynDisc	0.3625	18			0.1321	0.4661	0.2372	3.36E-04	0.2207	2.80E-03	0.1899	7.72E-03	0.2297	7.63E-06
	Best-Fit	0.3482	18	0.1246	0.5507			0.2267	1.40E-03	0.2102	7.97E-03	0.1997	0.01925	0.2538	1.14E-05
	DefDisc	0.2605	18	0.01952	0.9997	0.03003	0.9988			0.1081	0.7246	0.03153	0.9986	0.2387	2.67E-04
	MLE	0.2741	18	0.03604	0.9978	0.04655	0.9931	0.1486	0.2899			0.07357	0.9458	0.2538	1.14E-05
	StateDisc	0.3125	18	0.03979	0.9929	0.05706	0.9829	0.2252	1.68E-03	0.1832	0.05935			0.2568	3.81E-06
E,S|I	DynDisc	0.2833	18			0.1922	0.03327	0.2417	1.64E-04	0.2417	1.64E-04	0.2005	0.01736	0.2568	3.81E-06
	Best-Fit	0.2709	18	0.06456	0.97			0.2102	7.97E-03	0.2237	2.00E-03	0.1224	0.5717	0.2492	3.81E-05
	DefDisc	0.1819	18	0.01502	0.9999	0.04655	0.9931			0.1021	0.7789	0.03453	0.998	0.1727	0.1061
	MLE	0.1927	18	0.01502	0.9999	0.03303	0.9983	0.1547	0.2341			0.05105	0.9896	0.1922	0.03327
	StateDisc	0.2667	18	0.05631	0.9836	0.1344	0.4367	0.2222	2.35E-03	0.2057	0.01184			0.2282	1.53E-05
R,S|V	DynDisc	0.3	5			0.1273	0.5938	0.2364	0.09375	0.1273	0.5938	0.2	0.2188	0.2545	0.0625
	Best-Fit	0.3741	5	0.1455	0.5			0.2182	0.1563	0.1455	0.5	0.1818	0.3125	0.2545	0.0625
	DefDisc	0.2077	5	0.03636	0.9375	0.05455	0.9063			0.05455	0.9063	0.05455	0.9063	0.2727	0.03125
	MLE	0.3239	5	0.1455	0.5	0.1273	0.5938	0.2182	0.1563			0.2	0.2188	0.2364	0.09375
	StateDisc	0.2	5	0.07273	0.8438	0.09091	0.7813	0.2182	0.1563	0.07273	0.8438			0.2727	0.03125

Table A.1 | Summary of results for all algorithms and all categories. This table supports the claims and data summarized in Figure 3 and Figure 4 from the main text and Figure A.9 by summarizing comparative algorithm performance broken down by data category. Rows are grouped by data category as named in the first column. The next column lists algorithm names. The next two columns give the median correct classification rate (CCR) and the number of recordings (N). The next ten columns show either the effect size (r_{sdf}) or the p-value for a one-tailed Wilcoxon signed rank test of the hypothesis that CCR was greater for the

algorithm named on the row than the algorithm named on the column. The last two columns compare to chance performance which was 1/8 for orientation and 1/6 for size or contrast.

The data categories that we included in table A.1 have scientific interest, but by pooling data from related categories we can gain clarity about specific variables. In particular, by pooling across recording types but controlling the stimulus variable we can separate different aspects of population dynamics. For all pooled orientation categories we get a median correct classification rate (CCR) of CCR=0.2156 (greater than chance by Wilcoxon signed rank: r_{sdf}=0.218, p=0.488×10^{-4}). For contrast we get CCR=0.2875, (r_{sdf}=0.2494, p=5.73×10^{-11}). For size we get CCR=0.3167, (r_{sdf}=0.2405, p=2.73×10^{-12}). This shows that while orientation information is less prominent in V1 dynamics than size or contrast, there is enough to be detected. By contrast, when we try to classify orientation based on deflection we get CCR=0.1392 (does not pass Wilcoxon signed rank, r_{sdf}=0.1453, p=0.1879). This is despite the fact that we do find orientation tuning curves that are well defined on trial-averages and that the most and least preferred orientations do evoke reliable deflection responses (see Figure A.1). There is simply enough trial-to-trial variability among the other orientations that deflection does not inform about orientation well, but a dynamical perspective is less susceptible to this variability. Crucially for theories of neural coding we have shown that neurons do have more information than they pass down stream. Therefore if an experimentor wants to "listen in" on upstream populations they are better off using a model-based approach like dynamical discrimination than by using summative measures like deflection.

We can also categorize the data by recording type. There we see that inhibitory current dynamics lead to CCR=0.3167, excitatory current dynamics lead to CCR=0.2833, and Vm dynamics (spikes removed) lead to CCR=0.275. Interestingly, synaptic inhibition appears to have more stimulus related data about contrast and size than excitation (r_{sdf}=0.067, p=0.0279, N=33). This underscores our key result that synaptic excitation has more fine-grained information about orientation than synaptic inhibition, as only synaptic excitation passed a Wilcoxon Signed rank test for distinguishability for chance. However the dataset is small enough to limit our ability to make distinctions when controlling for both recording type and stimulus variable. When instead of comparing to chance we directly compared $\mathcal{E}_{\mathcal{O}}\mathcal{I}$ to $\mathcal{I}_{\mathcal{O}}\mathcal{I}$, it did not pass the Wilcoxon signed-rank test for distinguishability. Because we have only 7 examples, the variability overwhelmed the relatively large effect size (r_{sdf}=0.0875) and reduced confidence in the difference (p=0.1043). It is likely that a larger dataset is required to make definitive judgements about narrow cross-category comparisons such as this using dynamical discrimination alone. Nonetheless comparisons across algorithms and chance are confidently detected. Furthermore we can include information from other algorithms to support cross-category insights. For example when best-fit Ξ are used as a basis for discrimination we also find that for contrast and size synaptic inhibition leads to better discriminability than excitation (r_{sdf}=0.058, p=4.304×10^{-3}) and the same was found when using deflection based discrimination (r_{sdf}=0.058, p=0.0125). The conclusion that inhibition generally has more stimulus information regarding size and contrast seems robust and highlights the uniqueness of the finding that only a dynamical interpretation and only synaptic excitation can reveal fine grained orientation information.

Lastly it is important to check that our spiking proxy (deflection) can function as a stimulus detector in accordance with the role that computational neuroscience theorizes for individual neurons in V1 that exhibit stimulus tuning. The value of tuning of neural firing is often theorized to be the detection of a key type of stimulus, rather than as continuously informing about the properties of stimulus (Keller and Mrsic-Flogel, 2018). So we tested whether our methods could be used to say whether a stimulus is either the preferred stimulus or the least preferred stimulus when presented with one or the other. We used the same Ξ matrices obtained before from hold-one-out out-of-sample testing, but retained only those fitted to recordings coinciding with presentations of the most and least preferred stimuli. We retrained and retested random forest classifiers on these Ξ matrices. Next we did the same for deflection. This a discrimination test that is very natural for deflection, whereas discriminating between relatively fine variations in stimuli is something we expect deflection to be poor at and thus motivated the development of dynamical discrimination. We find that deflection exhibits high variability in cell-to-cell performance, but generally excels. The only category not greater than chance was membrane potential recordings coinciding with variations in size, which had only five cells, leading to poor statistical resolution despite a median CCR of 0.7. The data supporting claims of significance in Figure A.7 are presented in table A.2. These findings confirm that our definition of deflection is a valid basis of comparison. Thus our extension to finer levels of comparison with dynamical discrimination are useful. Dynamical discrimination had a lower upper limit on performance. Like deflection based discrimination it had only one category fail to be distinguishable from chance: inhibitory current recordings coinciding with variations in orientation. This was also the worst performing category in our general

findings. Deflection out performs dynamical discrimination at binary discrimination tasks, but this not surprising given it is a natural challenge for deflection and the Ξ matrices we use for dynamical discrimination were extensively optimized for a different task (fine distinctions rather than broad ones). Finally, the usefulness of deflection for binary classification underscores the insight that neurons receive more information than they pass on to other brain regions.

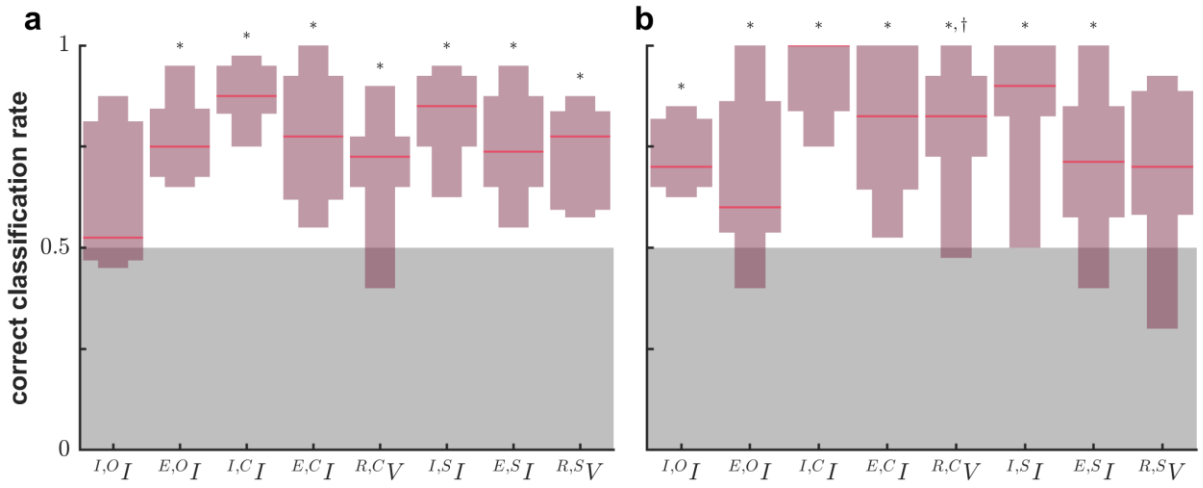

Figure A.1 | Deflection is useful as a binary discriminator in accordance with stimulus selectivity theory. a, Same as Figure 3b, except showing the classification ability of using dynamical discrimination when limiting to just the data from each cell's least and most preferred stimulus. Distinguishability from chance according to the Wilcoxon signed-rank test is indicated with *. **b,** Same as Figure 3a except showing classification ability of using deflection when limiting to just the data from each cell's least and most preferred stimulus. Distinguishability from chance is indicated with *, and from dynamical discrimination with †.

category	algorithm	median CCR	r_{sdf}	p-value	chance r_{sdf}	chance p-value
All	DynDisc	0.7875	0.091341	0.974027	0.249959	1.19E-30
	DefDisc	0.825	0.141937	0.02607	0.243192	1.6E-27
I,O$_I$	DynDisc	0.525	0.07619	0.851563	0.2	0.140625
	DefDisc	0.7	0.190476	0.1875	0.266667	0.007813
E,O$_I$	DynDisc	0.75	0.195238	0.164063	0.266667	0.007813
	DefDisc	0.6	0.071429	0.867188	0.233333	0.046875
I,C$_I$	DynDisc	0.875	0.058065	0.946045	0.258065	3.05E-05
	DefDisc	1	0.167742	0.056824	0.258065	3.05E-05
E,C$_I$	DynDisc	0.775	0.129032	0.505219	0.258065	3.05E-05
	DefDisc	0.825	0.129032	0.505219	0.258065	3.05E-05
R,C$_V$	DynDisc	0.725	0.014286	0.996094	0.252381	0.00293
	DefDisc	0.825	0.2	0.009766	0.257143	0.001953
I,S$_I$	DynDisc	0.85	0.074324	0.942467	0.256757	3.81E-06
	DefDisc	0.9	0.182432	0.060661	0.22973	7.63E-06
E,S$_I$	DynDisc	0.7375	0.114114	0.513832	0.256757	3.81E-06
	DefDisc	0.7125	0.115616	0.495377	0.248498	3.81E-05
R,S$_V$	DynDisc	0.775	0.172727	0.34375	0.272727	0.03125
	DefDisc	0.7	0.1	0.71875	0.227273	0.125

Table A.2 | binary discrimination data. This table gives the exact test values summarized in Figure A.1. It is the same as table A.1 except that only two algorithms are compared. Rows are grouped by data category, indicated in the first column, algorithms identified in the second. The third column contains the median correct classification rate (CCR). For the next two columns the Wilcoxon signed rank effect size (r_{sdf} and p-value) shown are for a test that the algorithm identified on the row outperformed the other algorithms in the same category. For example the All - DynDisc row in the r_{sdf} column is the effect size for a comparison to All - DefDisc. The last two columns show a comparison to chance (CCR=1/2).

A.2 Hyperparameter Optimization Reveals Epoch Dependence of Dynamical Discrimination

Dynamical discrimination based on genetic SINDy requires choosing a large number of hyperparameters. This is discussed in methods. Some of these choices are of scientific merit. For example, it may seem intuitive that the steady-state of stimulus response would

better represent dynamics because it avoids “transients” and would therefore permit better dynamical discrimination. However, we found that a brief period at the beginning of the stimulus response allowed the highest levels of dynamical discrimination, perhaps supporting the findings of fast attractor dynamics in vision (Miconi et al., 2016) and belying the implication of calling these periods “transients” (Müller et al., 2001; Mazor and Laurent, 2005; Bondanelli and Ostojic, 2018). Additionally, it is important from the standpoint of rigor and transparency that all publications which make extensive use of machine learning should report on the variations they tried and demonstrate that they optimized for a parameter that is independent of the scientific comparisons they see. Figures A.2-A.6 show a table of classification performance in various scenarios, including wether or not to use an ensemble of ${}^{B}\Xi$ matrices (see methods) for classification, the regularization factor which influences how many non-zero elements there are in the ${}^{B}\Xi$ examples, how many dimensions to include in the dimensionality expansion, whether to use second order differential equations, and finally what stimulus presentation epoch to use for dynamical discrimination. The data are from the cross-validation set where overfitting is also observed. However overfitting is stable, not contributing greatly to the variability (Figure 2b). The final holdout performance is lower but large differences in performance are preserved.

Discriminability is an experimental variable familiar to neuroscience which attempts to characterize the ability to classify a stimulus given a neural response and a theory about how information is encoded in it. It has been found that discriminability is highest at the early stages of a neural response, the on-epoch (Müller et al., 2001). Our analysis also

shows that discriminability is higher in the on-epoch versus the “off-epoch”, assuming our hypothesis about stimulus information modulating parameters of the governing dynamical system. This is true for all stimulus types and signal types. Importantly we also examined the epoch dependence of deflection and found that classification based on deflection is completely ineffective for the “off-epoch”. The fact that this analysis reproduces known details of neural computation lends credibility to the method and is novel in the sense that this has never been demonstrated at the single neuron level.

We also see that additional dimensions and second order derivatives are generally not required. We believe that this is because the first three dimensions represent mixtures of variables, not single variables. This was confirmed by running FitzHugh-Nagumo (Izhikevich and FitzHugh, 2006) single neuron simulations driven with sine-wave current injection. The original dynamics have four dimensions but dimensionality expansion did produce two dimensions corresponding to FitzHugh-Nagumo and another two dimensions corresponding to a sine wave. Consider the possibility that there are multiple sets of subpopulations in upstream neurons and each set is independent of the other sets and each set has its own unique dynamics. Our dimensionality expansion method (time delay and SVD) is not guaranteed to separate them such that each dimension corresponds to only one set of subpopulations. Therefore each dimension may represent a mixture of variables from independent systems. Thus, dimensionality expansion would give a maximally compact representation and a small number of dimensions would be needed, in our case three. In some cases the higher dimension systems allow some improvement in classification, but

not a large enough difference to be confident that it would carry into final holdout performance, or to justify the additional computational time required.

Lastly we also test the dimensionality reduction algorithm tried, SVD prioritizes orthogonality between the components it estimates. We tested it against independent component analysis (ICA) which prioritizes statistical independence instead. It was found that ICA did not perform better, though it did produce visually distinct trajectories with much faster dynamics.

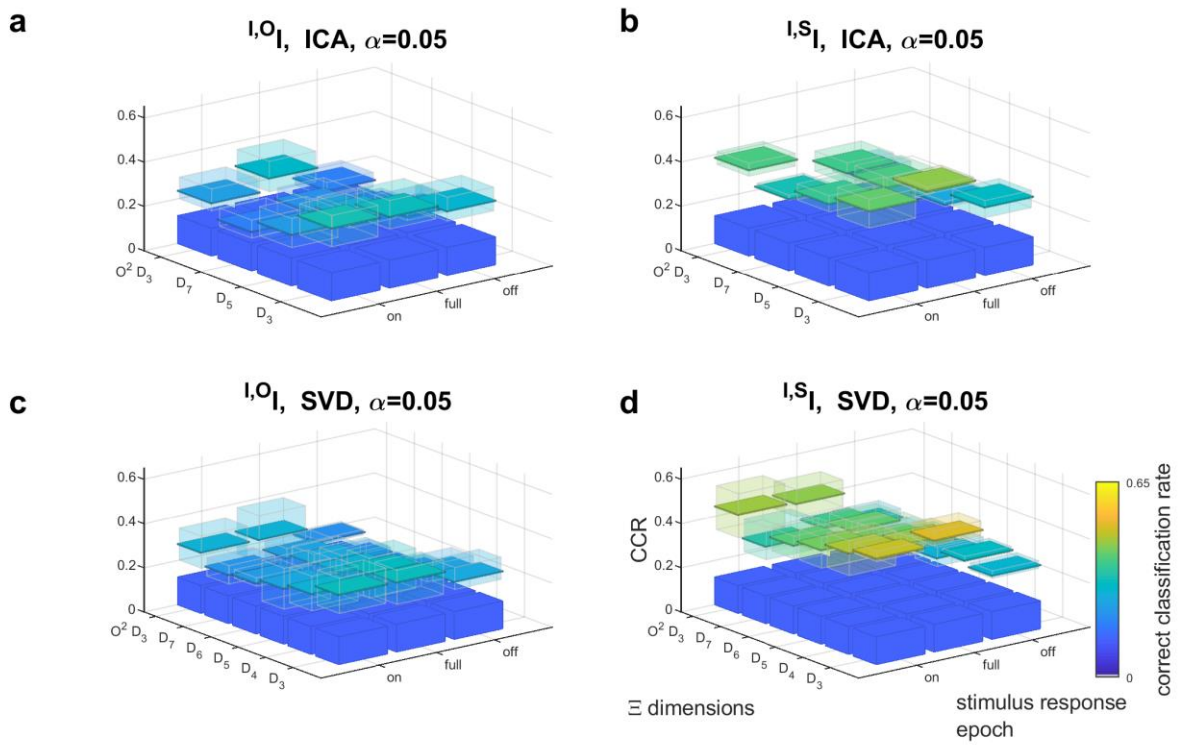

Figure A.2 | Comparisons for hyperparameter optimization part one: weak regularization, SVD compared with ICA along with epoch and dimension options. Four 3D barcharts showing the effect of several hyperparameters are plotted in a grid. It is evident that the off epoch is least informative and SVD outperforms ICA. **a,** Synaptic inhibition while orientation was varied was tested in combination with ICA based time-delay dimensionality expansion and a weak sparseness regularization. The vertical axis is the correct classification rate, the color also indicates the correct classification rate to aid visual comparison. Fully colored planes show median values. Translucent boxes show the variability (the min and max of the cells tested for that hyperparameter combination). The solid colored boxes show the rate of correct classification by chance. The horizontal axes are labeled, giving the hyperparameters tested for each bar position. There are three stimulus epochs and varying options for the dimensions (columns of Ξ) to include when fitting. The subscript of D denotes how many first order dimensions are kept "O^2" denotes that the fitted ODE model included second order derivatives, hence O^2 D_3 corresponds to a Ξ matrix with six columns. For each combination of hyperparameters the three cells with the greatest number of trials were tested. **b,** The same as in a except that synaptic inhibition was recorded while size varied. **c,** The same as in a except that SVD was used for time-delay based dimensionality expansion. **d,** The same as in c except that synaptic inhibition was recorded while size varied.

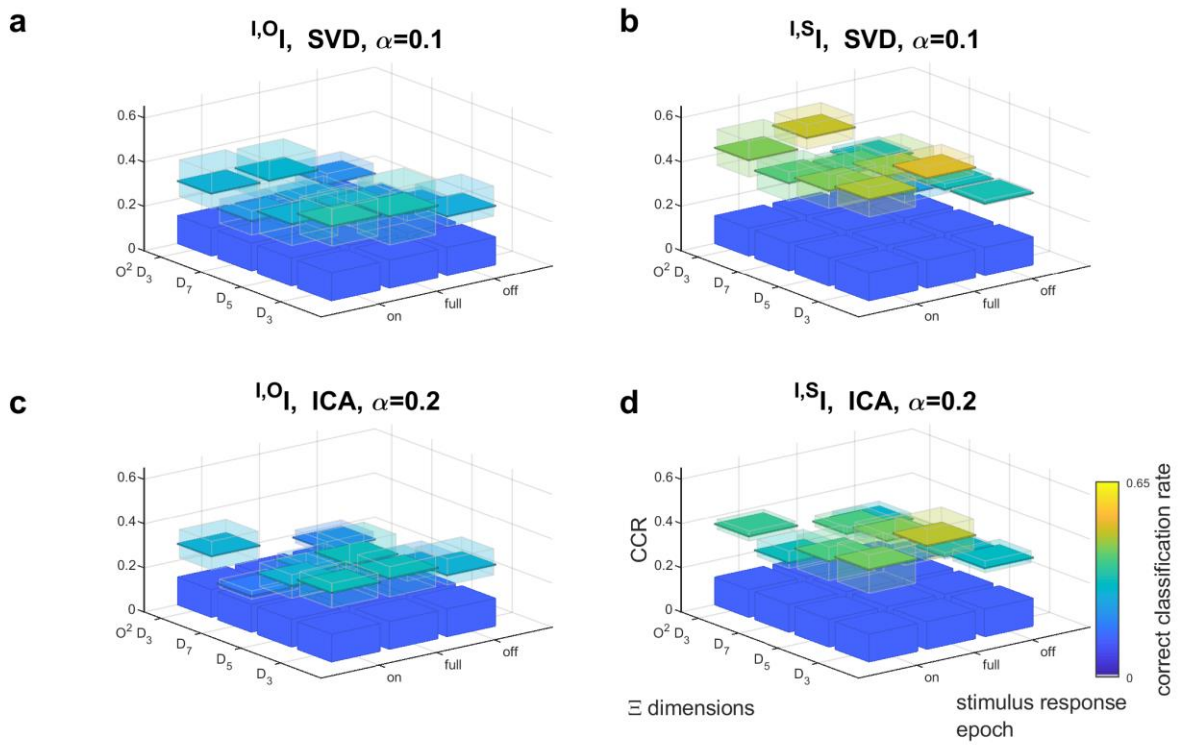

Figure A.3 | Comparisons for hyperparameter optimization part two: varied regularization and dimensionality reduction along with epoch and dimension options. Four 3D barcharts showing the effect of several hyperparameters are plotted in a grid. **a,** Synaptic inhibition while orientation was varied was tested in combination with SVD based time-delay dimensionality expansion and a weak sparseness regularization. The vertical axis is the correct classification rate, the color also indicates the correct classification rate to aid visual comparison. Fully colored planes show median values. Translucent boxes show the variability (the min and max of the cells tested for that hyperparameter combination). The solid colored boxes show the rate of correct classification by chance. The horizontal axes are labeled, giving the hyperparameters tested for each bar position. There are three stimulus epochs and varying options for the dimensions (columns of Ξ) to include when fitting. The subscript of D denotes how many first order dimensions are kept, while "O^2" denotes that the fitted ODE model included second order derivatives, hence O^2 D_3 corresponds to a Ξ matrix with six columns. For each combination of hyperparameters the three cells with the greatest number of trials were tested. Comparing to Figure A.2c, and A.4a reveals that increased sparseness regularization continues to improve performance. **b,** The same as in a except that synaptic inhibition was recorded while size varied. Comparing to Figure A.2b, and A.4b reveals that increased sparseness regularization continues to improve performance. **c,** The same as in a except that ICA was used for time-delay based dimensionality expansion, and sparseness regularization is moderate. Comparison with Figure A.4a shows that ICA continues to underperform SVD even with higher sparseness regularization. **d,** The same as in c except that synaptic inhibition was recorded while size varied. Comparison with Figure A.5b shows that ICA continues to underperform SVD even with higher sparseness regularization.

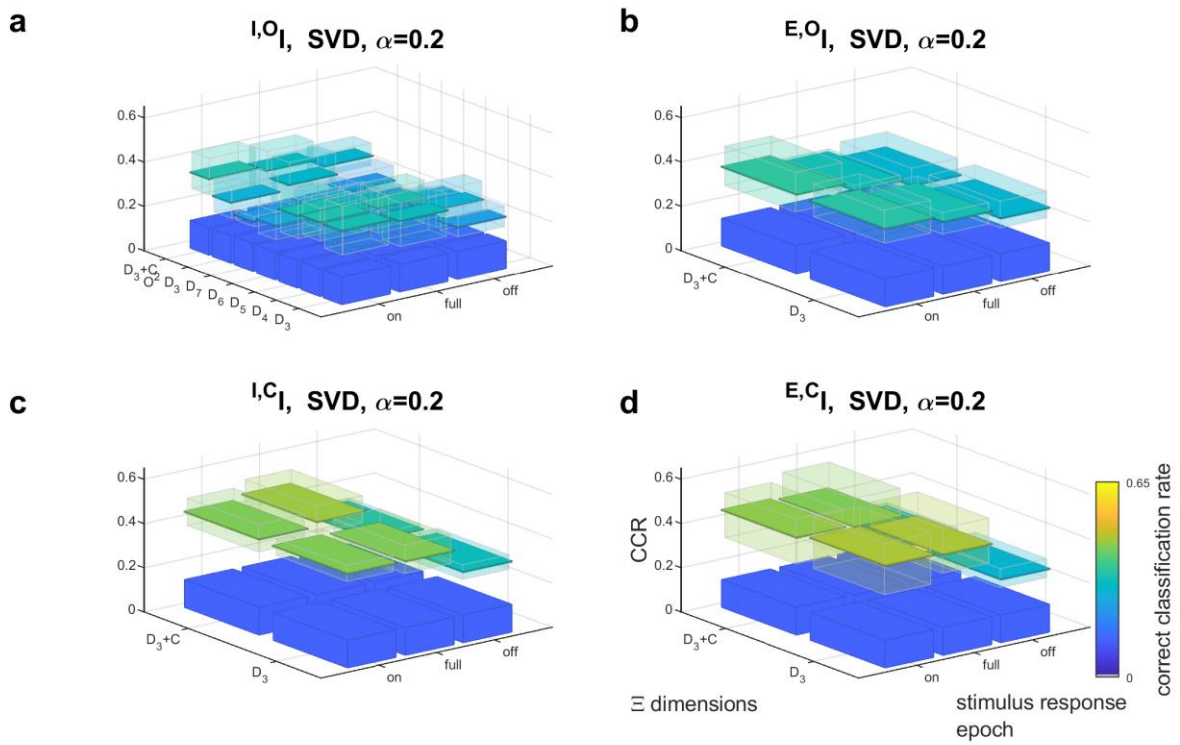

Figure A.4 | Comparisons for hyperparameter optimization part 3: moderate regularization, four categories showing epoch and dimension dependence. Four 3D barcharts showing the effect of several hyperparameters are plotted in a grid. **a,** Synaptic inhibition while orientation was varied was tested in combination with SVD based time-delay dimensionality expansion and a moderate sparseness regularization. The vertical axis is the correct classification rate, the color also indicates the correct classification rate to aid visual comparison. Fully colored planes show median values. Translucent boxes show the variability (the min and max of the cells tested for that hyperparameter combination). The solid colored boxes show the rate of correct classification by chance. The horizontal axes are labeled, giving the hyperparameters tested for each bar position. There are three stimulus epochs and varying options for the dimensions (columns of Ξ) to include when fitting. The subscript of D denotes how many first order dimensions are kept, while "O^2" denotes that the fitted ODE model included second order derivatives, hence O^2 D_3 corresponds to a Ξ matrix with six columns, and "+C" denotes that the fitted ODE model included a quenched noise term (four column Ξ). For each combination of hyperparameters the three cells with the greatest number of trials were tested. The pont D_3 & "on" was chosen for final analysis. The colormap evinces that no other dimension and epoch choices performed significantly better given the variability. The "off" epoch performed poorly. Comparison with Figure A.6a and A.6c shows that sparseness regularization has plateaued. **b,** The same as in a except that synaptic excitation was recorded while orientation varied and only two dimension options are tested. We can see that the inclusion of a driving noise term did not enable significantly better stimulus discriminability given the variability. **c,** The same as in b except that synaptic inhibition was recorded while contrast varied. **d,** The same as in b except that synaptic excitation was recorded while contrast varied.

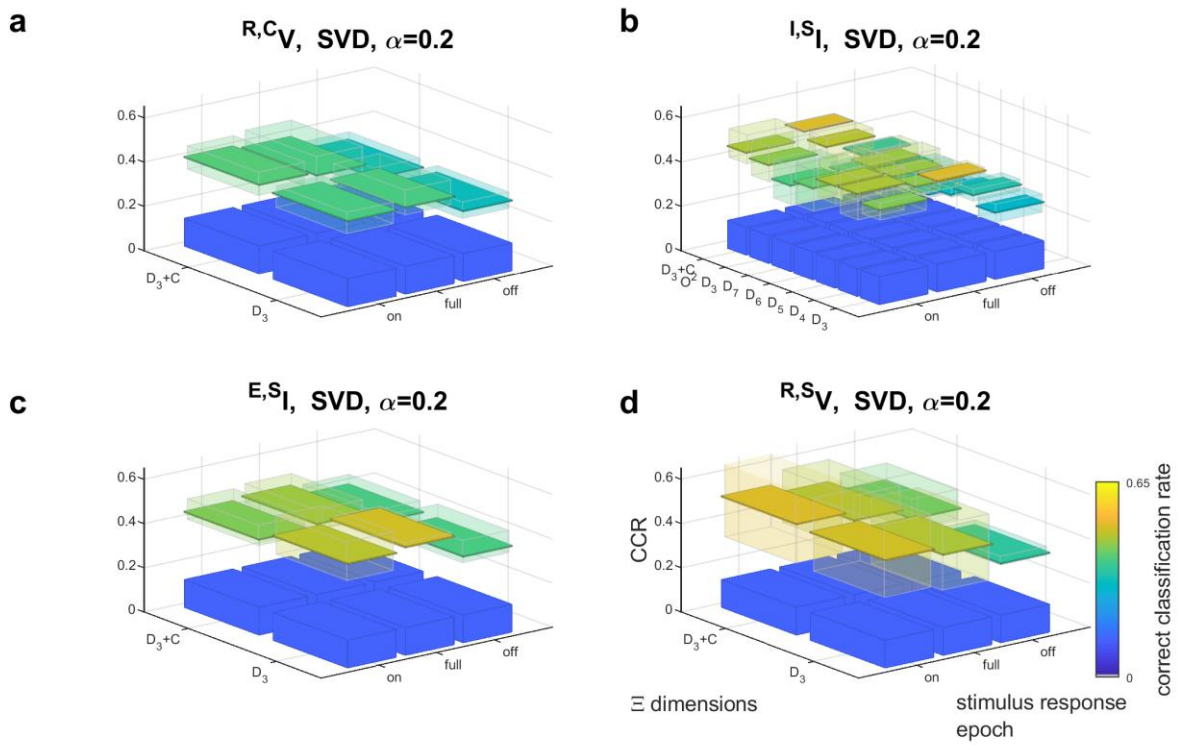

Figure A.5 | Comparisons for hyperparameter optimization part four: moderate regularization, four categories showing epoch and dimension dependence. Four 3D barcharts showing the effect of several hyperparameters are plotted in a grid. **a,** Membrane potential while contrast was varied was recorded and spikes were removed. The vertical axis is the correct classification rate, the color also indicates the correct classification rate to aid visual comparison. Fully colored planes show median values. Translucent boxes show the variability (the min and max of the cells tested for that hyperparameter combination). The solid colored boxes show the rate of correct classification by chance. The horizontal axes are labeled, giving the hyperparameters tested for each bar position. There are three stimulus epochs and varying options for the dimensions (columns of Ξ) to include when fitting. The subscript of D denotes how many first order dimensions are kept, while "+C" denotes that the fitted ODE model included a quenched noise term (four column Ξ). For each combination of hyperparameters the three cells with the greatest number of trials were tested. We can see that the inclusion of a noise term did not significantly improve classification given the variability. The pont D_3 & "on" was chosen for final analysis. **b,** The same as in a except that synaptic inhibition was recorded while size varied and many more dimension options were tested. The axis mark "O^2" denotes that the fitted ODE model included second order derivatives, hence O^2 D_3 corresponds to a Ξ matrix with six columns. The pont D_3 & "on" was chosen for final analysis. The colormap evinces that no other dimension choices performed better given the variability. The "off" epoch performed poorly, but the "full" epoch performed well. Comparing to the other plots including those on Figure A.4 shows that the "full" response epoch did not generally perform better given the variability. Comparison with Figure A.6b and A.6d shows that sparseness regularization has plateaued. **c,** The same as in a except that synaptic excitation was recorded while size varied. **d,** The same as in a except that membrane potential while size was varied was recorded and spikes were removed.

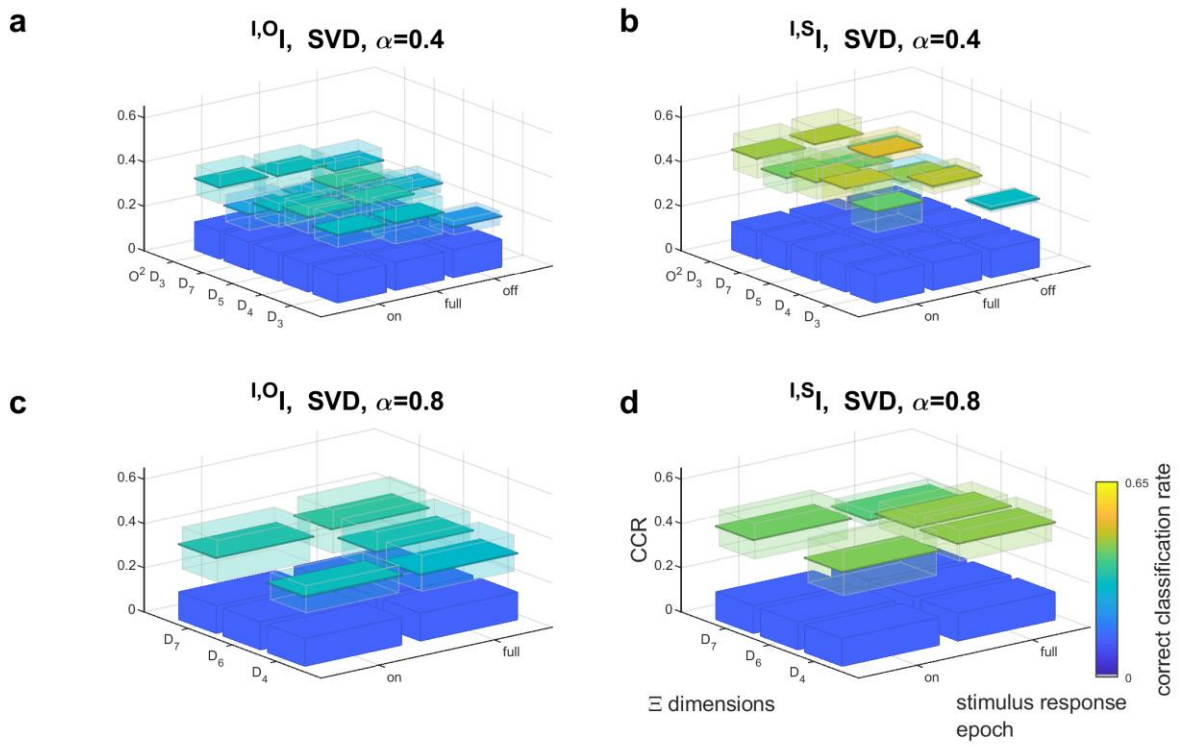

Figure A.6 | Comparisons for hyperparameter optimization part five, varied strong regularization along with epoch and dimension options. Four 3D barcharts showing the effect of several hyperparameters are plotted in a grid. **a,** Synaptic inhibition while orientation was varied was tested in combination with SVD based time-delay dimensionality expansion and strong sparseness regularization. The vertical axis is the correct classification rate, the color also indicates the correct classification rate to aid visual comparison. Fully colored planes show median values. Translucent boxes show the variability (the min and max of the cells tested for that hyperparameter combination). The solid colored boxes show the rate of correct classification by chance. The horizontal axes are labeled, giving the hyperparameters tested for each bar position. There are three stimulus epochs and varying options for the dimensions (columns of Ξ) to include when fitting. The subscript of D denotes how many first order dimensions are kept "O^2" denotes that the fitted ODE model included second order derivatives, hence O^2 D_3 corresponds to a Ξ matrix with six columns. For each combination of hyperparameters the three cells with the greatest number of trials were tested. Through comparison to Figure A.2c, A.3a, A.4a, and panel c here, we see that the impact of sparseness regularization is minimal beyond about α=0.2. **b,** The same as in a except that synaptic inhibition was recorded while size varied. Through comparison to Figure A.2d, A.3b, A.4b, and panel d here, we see that the impact of sparseness regularization is minimal beyond about α=0.2. **c,** The same as in a except that sparseness regularization is now α=0.8. Data was not collected for D_6 and "on", or any "off" epochs. **d,** The same as in c except that synaptic inhibition was recorded while size varied.

A.3 Analysis of Dynamical Stability Underscores an Attractor Dynamics Interpretation

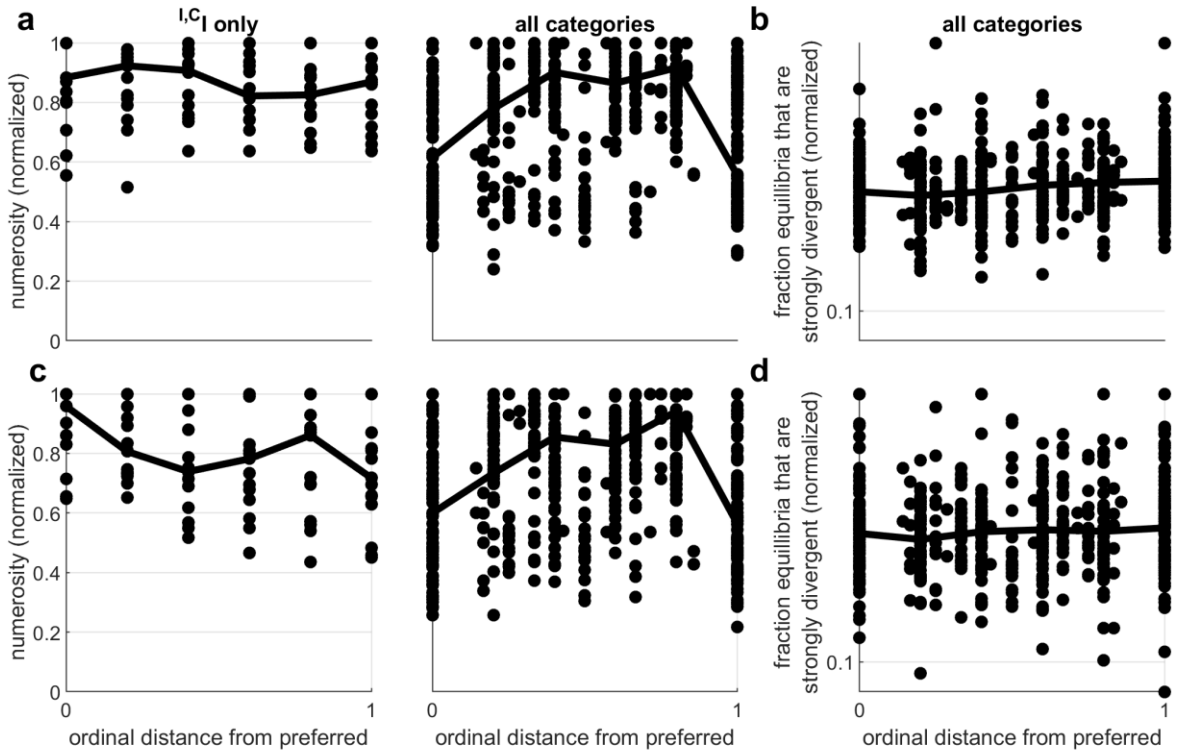

Figure A.7 | Scatter plots summarizing differences between stimulus conditions according to various indicators of bifurcations. a, The first row of plots (a and b) shows data from analyzing the fitted ODE which modeled the dynamics the best (best-fit Ξ). The horizontal axis is the ordinal distance from the most preferred stimulus. The vertical axis plots fixed point numerosity. The solid black line shows the median of 6 deciles of ordinal distance. The left column shows a single category ^{I,C}I, which is not U-shaped and no significant trend is present.The right column shows the result of pooling all the data, a U-shaped trend is visible but upon further scrutiny it is due to combining data from cells with opposite trends, not a fundamental U-shaped trend for individual cells. **b,** Data from the Ξ matrices (ODE models) that best describe the trajectories are plotted. The vertical axis plots the fraction of fixed points (equilibria) whose L^2 norm of negative real-valued parts was smaller than the L^2 norm of positive real-valued parts. A trend is present but weak. **c,** Same as in a except showing results from the fitted ODE that permited best classification (dynamical discrimination). A significant trend is present for the data from ^{I,C}I. Many cells had the opposite trend (not always significant). Consequentially pooling the data (plotted on the right) shows an inverted U. After aligning cells with opposite trends an overall trend is significant. **d,** Same as in b except showing results from the fitted ODE that permited best classification (dynamical discrimination) . No trend is detected.

Figure A.7 shows the results of analysis of Ξ matrices pursuant to stability and bifurcation analysis from nonlinear dynamics. Real valued fixed points are found by solving for them (equating columns of Ξ to zero) with MATLAB's symsolve computer algebra system. There

is no guarantee that any real-valued fixed points exist, as the system may be overdetermined or underdetermined. However this is rare with these data. The best-fit Ξ matrices yield real-valued fixed points for 99.2% of recordings, and the Ξ matrices from dynamical discrimination yield real-valued fixed points for 87.8% of recordings. In Figure A.7, (left and center columns) we see that the number of real-valued fixed points varies based on the stimulus coinciding with the data Ξ was fitted to. If the number of fixed points change when varying Ξ parameters then a bifurcation is identified. Thus we have additional evidence for distinct stimulus-evoked dynamics. The category of inhibitory current recordings co-occurring with varied contrast shows the most consistent cell-to-cell pattern. There is a trend toward fewer fixed points as the presented stimuli gets further from the preferred stimuli when looking at ODEs that permit the best classification performance (Spearman correlation $r=-0.2372$, $p=0.0244$, Figure A.7c), but not for ODEs that best model dynamics ($r=-0.1520$, $p=0.1528$, Figure A.7a).

We also find differences in the convergence/divergence of behavior near fixed points. When linearized near a fixed point (where derivatives are very small) the behavior of the maximum eigenvalue governs the stability of the dynamics near that fixed point. If multiple fixed points exist we linearize around each one. If the real value of the maximum eigenvalue passes through zero when changing Ξ parameters then a different kind of bifurcation in the dynamics is identified. We plot a related concept "net convergence", in the right most panel of Figure A.7 and show that it weakly depends on Ξ matrices. If the three eigenvalues of the Jacobian-linearization at each fixed point have both positive and negative real components and the L^2 norm of the negative real-valued components is larger

than the L^2 norm of the positive real-valued components then the fixed point is "net-convergent". The slight trend is that the fraction of fixed points that are net divergent tends to increase for stimuli a larger ordinal distance from the preferred stimulus. It is not necessary to control cell-to-cell variability to observe this effect. It is most detectable for best fit ODEs (Spearman Correlation, $r=0.1093$, $p=0.0075$) and not significant for ODEs that permit the best classification ($r=0.0717$, $p=0.0800$). This weak trend is double checked by performing the Wilcoxon rank-sum test to compare the median fractions of the data above and below the 25th and 75th percentiles of normalized ordinal distance to preferred stimulus respectively, ($r_{sdf}=0.1506$, $p=0.0049$ for best fit ODEs and $r_{sdf}=0.1015$, $p=0.0419$ for ODEs that permit the best classification).

An important consideration for attempting to identify bifurcations by using linear stability analysis is cell-to-cell variability. While one cell may have more fixed points for the least preferred stimuli than for the most preferred there is no a priori reason why a different cell cannot show the opposite trend, or even have the same number of fixed points for the most and least preferred stimuli but a different number for the intermediate stimuli. This effect is seen in our data, and makes it difficult to gather population statistics. Consider the U-shaped trend in the central column of Figure A.7 (A.7a and A.7c). In the first case the U-shape trend exists for individual cells as evinced by a high correlation between reliability and distance from the most or least preferred stimulus (which takes the plot and "folds it vertically" at an individual level). That correlation showed that there is a non-monotonic trend. There is no such correlation for N. When all data from all cells are normalized without aligning trends and plotted against the normalized ordinal distance from only the

preferred stimulus there is an inverted U appearance. This inverted U-shaped trend is not a coincidence, as revealed by affecting the following change of variables on the horizontal axis $x = |x - 0.5|$ and then obtaining the Spearman correlation (r=-0.2414, p=2.248×10^{-9} for best fit ODEs and r=-0.2171, p=8.258×10^{-8} for ODEs that permit the best classification). This is similar to but not the same as the ordinal distance from either the most or least preferred stimulus. Nonetheless this inverted U-shape is due to the fact that some cells have a positive trend while others show a negative trend and not because cell's individually have an inverted U-shaped trend. The existence of two strong but opposite trends among different cells is shown to be the cause of the inverted U-shaped trend via the following analysis. For each cell we obtain the average number of fixed points for the ODEs fitted to each stimulus, denoted by $N_{i,j}$ where i denotes the cell index and j denotes the stimulus index. Then we subtract out the mean, multiply by the sign of the value at the preferred stimulus (which aligns the trends) and add the mean back,$N'_{i,j} = (N_{i,p}/|N_{i,p}|) \cdot (N_{i,j} - \langle N_{i,j}\rangle_j) + \langle N_{i,j}\rangle_j$, where p denotes the index of the preferred stimulus. Finally, we normalize the largest value to one $N''_{i,j} = N'_{i,j}/\max_j(N'_{i,j})$. Then we measure the correlation of these aligned trends and find a Spearman correlation of (r=-0.1907, p=2.658×10^{-6} for best fit ODEs and r=-0.2249, p=2.721×10^{-8} for ODEs that permit the best classification). Thus the aligned trends explain the correlation with distance from the center of the overall inverted U, and consequently evinces a bifurcation-type effect because of the number of fixed points changes. For some cells the preferred stimulus has fewer fixed points while for other cells it has more. Cell-to-cell variability is not the only factor, a similar factor is Ξ to Ξ variability. We independently train 45 different Ξ matrices for each cell and let the Ξ matrices "vote" on a correct classification. Hence, each Ξ matrix is a

different model of dynamics, akin to a different "perspective". Hence different Ξ matrices also show different trends. For this reason we report on only the Ξ matrix which had the most optimal objective function value at the last generation of the genetic algorithm.

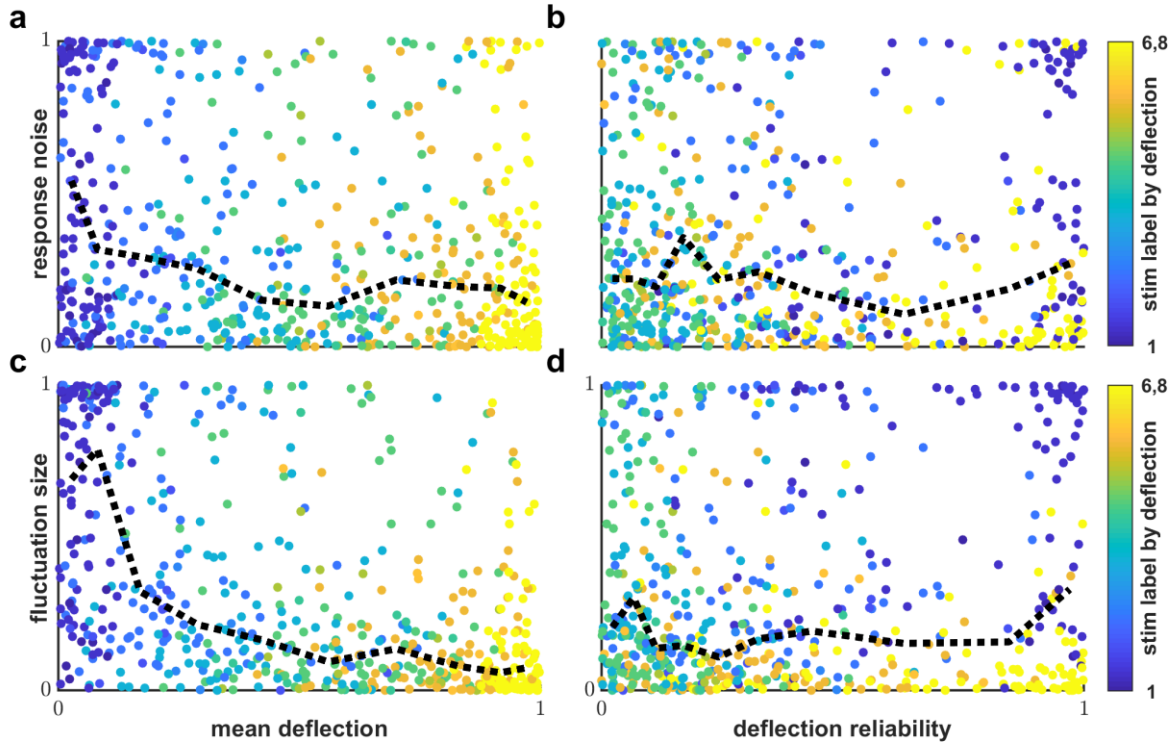

Figure A.8 | Scatter plots showing how measures of noise and fluctuation size relate to deflection and reliability. a, A normalized measure of noise, the mean residual error after subtracting mean stimulus-dependent response, is plotted against normalized mean deflection. Colors indicate stimulus labels according to deflection rank (color bar right of panels b and d). The black dashed line indicates central tendency: the median noise in 10 deciles of mean deflection. Unlike mean reliability, which showed a strong U-shaped trend (high reliability for extremes of deflection), noise decays with increasing deflection. **b,** Normalized noise is plotted against reliability. Colors are the same as in a, the dashed line are the median noise for 10 deciles of reliability. If reductions in noise alone caused greater reliability we would expect to see a trend, and do not. **c,** A normalized measure of fluctuation size, the coefficient of variation for time-series points within the recording snippet, is plotted against normalized mean deflection. The pattern is very similar to that found in noise (panel a). Neither noise nor fluctuation size recapitulate the U-shape but it would be possible that they worked together if they showed opposite trends (e.g. low deflection high-reliability is due to small fluctuation while high-deflection high-reliability is due to low noise). However, both fluctuation size and noise are highest for low-deflection evoking stimuli therefore they do not work in combination. **d,** Normalized fluctuation size is plotted against normalized reliability. No correlation is found.

In the main text we juxtaposed our limited findings about linear stability analysis of ODE fixed points with our finding that extremes of deflection evoked the most reliable deflections on average. This has a possible explanation in dynamics, whereby stimuli that evoke intermediate deflections do so because the dynamics they evoke is inherently less repeatable (e.g. sensitive dependence on initial conditions) or because they sometimes evoke the dynamics associated with the lowest deflection values and at other times they evoke the dynamics associated with the highest deflection values. Greater reliability would be due to attractors following very similar trajectories. The trajectories themselves could cover any range of values and start at any point in their paths. Thus one attractor could produce trajectories that display larger apparent fluctuations than another attractor, yet a path integration measure (e.g. deflection) on two of these trajectories would produce similar values (i.e. high reliability). Alternatively two trajectories from the same attracting set could have high relative error because they start at different points (e.g. phase-shifted sinusoids), yet again in certain situations integration of the trajectory would yield similar values. A counter hypothesis is that the reliability is simply due to quenched variability in a random process that has similar governing dynamics in any situation. In this paradigm greater reliability would arise because either fluctuations are small or noise (relative error) is small. Hence we measured fluctuation size and relative error (see methods). If either alone explains the reliability patterns we saw then they would anti correlate strongly with reliability. Neither did (see Figure A.8b, A.8d), our measure of fluctuation size (coefficient of variation), gave a Spearman correlation of $r=-0.013$, $p=0.726$, and our measure of noise gave $r=-0.038$, $p=0.317$. Noise and fluctuation size could work together to produce the reliability pattern if one measure was very low for small deflections and the

other was very small for large deflections. This was not seen either, both strongly anticorrelated with deflection, our measure of fluctuation size (coefficient of variation), gave a Spearman correlation of r=-0.5, $p=8.179\times10^{-45}$, and our measure of noise gave r=-215, $p=1.252\times10^{-8}$. Because both measures strongly anti-correlated with deflection, instead of having a U-shaped trend or opposite trends, and because neither anticorrelated with reliability than an understanding founded on random-process origins would predict low reliability for small deflections and this was notably contradicted.

A.4 Maximum Likelihood Estimation of Stimulus Reveals Dimensionality Expansion is Not Sufficient

One way to test whether dimensionality expansion alone (without dynamical systems) is enough without the dynamical systems perspective is to simply use the probability of a trajectory being limited to a region of state space. By plotting the carefully selected example trajectories seen in Figure 2c we can assess whether all the information gleaned from dynamical discrimination is evinced more simply as the confinement of trajectories to regions of phase space, such as one conic surface nested inside another, or oscillating around centers that are displaced from one another. It is natural to describe these shapes with cylindrical coordinates. Although these patterns are visible, like deflection, they are highly variable and must be carefully assessed.

To capture the effects of dimensionality expansion without an ODE-based classifier we use a general method for trajectory classification, maximum likelihood estimation. This method acknowledges that no trajectory explores all of the available state space and that if any

classification is possible then it must be that some regions of state space are more likely to be explored than others. So it uses 75% of the recordings from one cell to learn the probability that a given point in state space will be occupied for each stimulus. For the remaining 25% of recordings each is assigned a probability that it co-occurred with each stimulus. The stimulus with the highest probability of co-occurrence is selected to be the prediction. As with dynamical discrimination, hold-out cross validation is repeated 510 times to get the average performance reported as final and presented in Figure A.9. Because the trajectories took the form of orbits confined to conic or cylindrical regions we ignore the angular dimension and just using the radial and axial coordinates to create the probability density functions. This defines regions of state space that are annuli with rectangular radial cross sections.

The performance of this approach was roughly comparable to deflection based classification with overall classification rates exceeding chance (r_{sdf}=0.225, p=2.04×10^{-15}) and categories ^{LC}I (r_{sdf}=0.258, p=3.05×10^{-5}), ^{EC}I (r_{sdf}=0.207, p=0.0206), ^{RC}V (r_{sdf}=0.2476, p=0.0049), ^{LS}I (r_{sdf}=0.2538, p=1.14×10^{-5}), and ^{ES}I (r_{sdf}=0.1922, p=0.0333) distinguishable from random chance (see Figure A.9c). This tells us that dimensionality expansion alone can reveal only a limited amount of additional information, and we get no additional information about orientation. This is likely due to the fact that regions of high density in the probability density maps overlap as seen in Figure A.9b, thus dimensionality expanded trajectories often have distinctive features that may be stimulus related they are not confined to easily separated regions of state space.

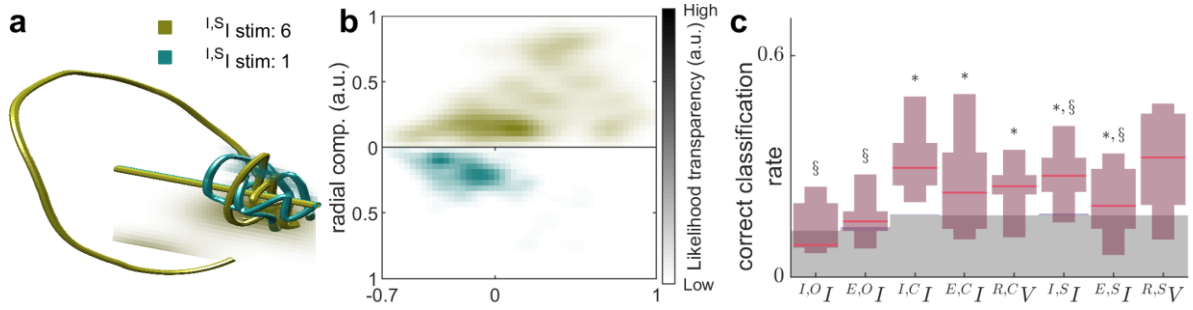

Figure A.9 | Trajectories vary by stimulus and occupy different but non-separable regions of state space a, A close view of two example trajectories. Aqua is an example of the least preferred stimulus and gold denotes the most preferred stimulus for the same cell. Synaptic inhibition was recorded while size varied. The reduced axial-radial cross section is shown as a shaded plane amidst the trajectories. This plane is most opaque in regions where the color-matched trajectory (aqua top, gold bottom) had a high likelihood of intersecting any given axial-radial cross section, this is known as the single trial axial-radial probability density map $M_t(z,r;i)$ (z is axial, r is radial, i denotes the trial index). **b,** The stimulus-dependent axial-radial probability density maps for the same two stimuli shown in b. Aqua (bottom) is $M_s(z,r;1)$, gold (top) is $M_s(z,r;6)$. This is created using all the trials coinciding with the selected stimuli. The distinctly conic region is apparent, as is the stimulus dependent nesting effect and the stimulus dependent axially translation for regions of maximal density. Selecting any single trial probability density map and selecting any stimulus-dependent probability map allows one to compute the likelihood that the single trial coincided with the selected stimulus. The method of Maximum Likelihood Estimation (MLE) amounts to trying all stimuli and finding the one with the highest likelihood. **c,** Same as in Figure 3a except it is showing the classification results for MLE trajectory classification. Greater than chance performance is indicated with *, and § indicates significantly worse performance than dynamical discrimination (Figure 3d). MLE is not better than deflection, indicating that the distinctive features seen in Figure 2g do not make trajectories separable.

References

Adesnik H (2017) Synaptic Mechanisms of Feature Coding in the Visual Cortex of Awake Mice. Neuron 95:1147–1159.e4.

Bondanelli G, Ostojic S (2018) Coding with transient trajectories in recurrent neural networks. arXiv [q-bioNC] Available at: http://arxiv.org/abs/1811.07592.

Breiman L (2001) Random Forests. Mach Learn 45:5–32.

Butts DA, Goldman MS (2006) Tuning curves, neuronal variability, and sensory coding. PLoS Biol 4:e92.

Carandini M, Ferster D (2000) Membrane potential and firing rate in cat primary visual cortex. J Neurosci 20:470–484.

Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV (2012) Neural population dynamics during reaching. Nature 487:51–56.

Ferster D (1986) Orientation selectivity of synaptic potentials in neurons of cat primary visual cortex. J Neurosci 6:1284–1301.

Gallego JA, Perich MG, Miller LE, Solla SA (2017) Neural Manifolds for the Control of Movement. Neuron 94:978–984.

Hough PVC (1959) Machine analysis of bubble chamber pictures. In: Conf. Proc., pp 554–558. inspirehep.net.

Izhikevich EM, FitzHugh R (2006) Fitzhugh-nagumo model. Scholarpedia J 1:1349.

Kaiser E, Kutz JN, Brunton SL (2018) Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. Proc Math Phys Eng Sci 474:20180335.

Kawakami K (2008) Supervised sequence labelling with recurrent neural networks. Ph D dissertation, PhD thesis Ph D Available at: https://pdfs.semanticscholar.org/a97b/5db17acc731ef67321832dbbaf5766153135.pdf.

Keller GB, Mrsic-Flogel TD (2018) Predictive Processing: A Canonical Cortical Computation. Neuron 100:424–435.

Kerby DS (2014) The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. Comprehensive Psychology 3:11.IT.3.1.

Laurent G (2002) Olfactory network dynamics and the coding of multidimensional signals. Nat Rev Neurosci 3:884–895.

Lee J-G, Han J, Li X, Gonzalez H (2008) TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. Proceedings VLDB Endowment 1:1081–1094.

Maharaj EA, D'Urso P, Caiado J (2019) Time Series Clustering and Classification. Chapman and Hall/CRC.

Mazor O, Laurent G (2005) Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. Neuron 48:661–673.

Miconi T, McKinstry JL, Edelman GM (2016) Spontaneous emergence of fast attractor dynamics in a model of developing primary visual cortex. Nat Commun 7:13208.

Müller JR, Metha AB, Krauskopf J, Lennie P (2001) Information conveyed by onset transients in responses of striate cortical neurons. J Neurosci 21:6978–6990.

Pandarinath C, O'Shea DJ, Collins J, Jozefowicz R, Stavisky SD, Kao JC, Trautmann EM, Kaufman MT, Ryu SI, Hochberg LR, Henderson JM, Shenoy KV, Abbott LF, Sussillo D (2018) Inferring single-trial neural population dynamics using sequential auto-encoders. Nat Methods 15:805–815.

Saha D, Leong K, Li C, Peterson S, Siegel G, Raman B (2013) A spatiotemporal coding mechanism for background-invariant odor recognition. Nat Neurosci 16:1830–1839.

Sauer T, Yorke JA, Casdagli M (1991) Embedology. J Stat Phys 65:579–616.

Strogatz SH (2018) Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. CRC Press.

Wilcoxon F (1992) Individual Comparisons by Ranking Methods. In: Breakthroughs in Statistics: Methodology and Distribution (Kotz S, Johnson NL, eds), pp 196–202. New York, NY: Springer New York.

Wolfe J, Houweling AR, Brecht M (2010) Sparse and powerful cortical spikes. Curr Opin Neurobiol 20:306–312.