Title
Contexts for the Emergence of Significant Neural Dynamics

Author
St. Clair, William Benjamin

Publication Date
2016

Peer reviewed|Thesis/dissertation

University of California, Merced

Dissertation

Contexts for the Emergence of Significant Neural Dynamics

Cognitive and Information Sciences

Author:

William Benjamin St. Clair

Cognitive and Information Sciences

wst.clair@ucmerced.edu

Committee:

David C. Noelle

Michael Spivey

Jeffrey Yoshimi

2016

Signatures of Approval:

Michael Spivey

Jeffrey Yoshimi

David Noelle, Committee Chair

Contents

Acknowledgements:

The Cognitive and Information Sciences Graduate Group for both its financial support and academic influence throughout the completion of this degree.

My committee—David Noelle, Jeffrey Yoshimi, and Michael Spivey—for their academic guidance and support throughout the years.

The members of the Computational Cognitive Neuroscience Lab (CCNL) for their feedback and resonance throughout this project. Particularly, Jeff Rodny for his feedback on simulation code, and Angelo Kyrilov, for his assistance with data management code.

CIS affiliated faculty, including, but not limited to: Chris Kello, Evan Heit, Rick Dale, Teenie Matlock, and Ramesh Balasubramaniam.

And, the many CIS graduate students and postdocs with whom I have had endless conversations that have shaped the perspectives essential to this dissertation. Including Jesse Edelstein, Morgan Flemming, Bryan Kerster, Tim Shea, Till Bergman, Drew Abney, Bodo Winter, and Marcus Perlman.

1 Introduction

How may neural dynamics carry conceptual information? This problem is made complex from many sources. The study of neural systems involves the integration of investigations across many physical scales, all of which have been argued to contribute to the understanding of how the brain processes information (genetics, neuron function, synapse function, neuroanatomy, neural networks, connectomics or whole brain connectivity, behavior, phenomenology). As a result of this, simply integrating implementational knowledge of the brain represents a large challenge by itself (Gordon, 2000). While complex in implementation, the dynamics of the brain may also come to be simple. Distinct experiences may consistently converge on similar responses or behaviors, providing the tantalizing possibility that a computationally meaningful reduction may emerge from the brains dynamics.

The appearance of unification, however, cannot be taken for granted. The emergence of meaningful dynamics, by itself, suggests an abundance of philosophical pitfalls that accompany its discussion. With this in mind, emergence in dynamical systems is here considered with attention to intertheoretic reduction and relationships across phenomenal scales, building to a treatment of emergence in dynamical systems neuroscience. This assessment motivates a formal notion of causal and existential meaning, in order to help contribute to our understanding of the neural code by relating its dynamics to the systems which underly it. Implications from this work are then applied to topics in computational cognitive neuroscience and philosophy of mind. Acknowledgement of the *embeddedness* of the significance of neural dynamics then motivates an extensive parametric study of the relationship between particular neural network structures and the emergence of significant neural dynamics.

In Chapter 2, I review necessary background material in dynamical systems neuroscience needed to motivate and understand later chapters. In Chapter 3, I review emergence in the philosophy of science literature (Section 3.1), show the relationship between the mystery of emergence and historico-theorical context (Section 3.2), review emergence in dynamical systems for intertheoretic reduction (Section 3.3), introduce a mathematical formalism for intertheoretic reduction and apply it to gliders in Conway's Game of Life (Sections 3.4 and 3.5), and, ultimately, apply the presented intertheoretic reduction formalism to emergence in neural systems (Section 3.8). Using this foundation, in Chapter 4, I consider implications of polychronous neuronal groups (PNGs) when they are viewed as propagating conceptual information. These implications include moving from conceptual vector spaces to conceptual PNGs (Section 4.1), capturing conceptual transitions by formalizing PNG trigger dynamics

in terms of existential and causal meaning (Section 4.2), characterizing conceptual blends (Section 4.3) through an emergent PNG space (Section 4.4), describe neural dynamics in prefrontal cortex (Sections 4.5.1–4.5.4), interpret frequent spiking dynamics in terms of PNGs (Sections 4.5.5–4.5.6), and use these intuitions to constrain a PNG account of linguistics, iconicity, and semantic meaning (Section 4.6).

The dependence of neural firing patterns on network topology is examined a large parametric analysis in Chapter 5, where 15,625 network architectures are analyzed to assess the relationship between different aspects of network topology and the firing patterns that emerge during 1000 milliseconds of tonic stimulation. Networks with too short of axonal propagation delays between cortical nuclei or with too little variability in propagation delay times, are found to exhibit unrealistic saturation levels, immediately achieving uniform spiking at 1000 Hz for dozens of milliseconds before collapsing to silence. Networks with longer and more variable delays are found to withstand more tonic input without "exploding," while also supporting a larger number of overall connections. Results from Chapter 5 suggest that larger and more variable delays may allow for greater connectivity between disparate regions without excessively endangering modularity or producing networks that consistently degenerate into epileptic-like patterns of neural firing. In Chapter 6, data from Chapter 5 is mined to identify networks that are significantly active from tonic rate coded stimulation without leading to informationally catastrophic explosions (Section 6.1). These networks are then analyzed in terms of the extent to which rate coded stimulation results in rate coded network dynamics (Sections 6.3–6.4). Results from Chapter 6 imply that the rate coded stimulation of a network does not guarantee rate coded dynamics downstream, supporting a characterization of frequent spiking as constraining the emergence of downstream spiking patterns without necessarily prescribing the information-bearing structure of that emergence.

2 Background

2.1 Polychronous Computation and Spiking Neural Networks

Polychronous Neuronal Groups (PNGs) have been proposed as a possible unit of representation in the brain (E. Izhikevich, 2006). A PNG is a reproducible, time-locked, spatiotemporal spike-timing pattern over a collection of neurons. They are reproducible in the sense that the sequence of spike times tends to replay when the input conditions experienced by the neuronal network are repeated. They are time-

locked in the sense that, once the PNG begins, the times between the spikes within the pattern are the same whenever the PNG is triggered. They are spatiotemporal in the sense that they are defined in terms of a specific set of neurons that participate in the pattern (spatial) as well as the precise times at which spikes occur (temporal). Since these spikes can be distributed across many times, in contrast to being synchronous, they are described as being *polychronous*. PNGs spontaneously emerge in spiking neuronal networks that incorporate variance in the amount of time it takes for an action potential to reach its receiving neurons (conduction delays). There is a growing literature on cognitive processing using PNGs (E. Izhikevich, 2007; Chorley & Seth, 2011; Szatmáry & Izhikevich, 2011), which builds on earlier work with *synfire chains* (Bienenstock, 1995; Hayon, Abeles, & Lehmann, 2005; Trengove, van Leeuwen, & Diesmann, 2013).

To understand the information-bearing properties of PNGs, it is important to understand how they are generated and propagated. An individual neuron remains at its resting potential until it receives, or "observes", a sufficient number of spikes in a short enough period of time, at which point this coincident input causes the neuron to generate an action potential of its own. This action potential is then, in turn, observed by the neurons to which this neuron projects. However, since it takes time for action potentials to propagate down axonal connections, there is a delay between when a spike is generated and when it is received (Swadlow, 1985, 1988). For example, in the cat brain, this delay can be as short as 0.1 ms, or as long as 44 ms (Swadlow, 1992). Since a cortical neuron may project to anywhere between 1,000 and 10,000 other neurons, a single action potential will be received at many different times. Thus, spikes that are synchronized on generation will not necessarily be synchronized on their receipt.

Typically, a single input spike is insufficient to drive the receiving neuron to fire an action potential, and the membrane potential of such a neuron is constantly decaying toward its resting potential. Within just a few milliseconds after receiving a single spike, the membrane potential of a neuron will return to its equilibrium state, removing the electrical effects of the spike (Cessac, Paugam-Moisy, Viéville, & et al, 2010). This highlights the need for synchrony in the arrival of spikes to initiate firing, but it is important to remember that spikes that are synchronized at the time of receipt will not necessarily be synchronized at the time of their initiation, due to variance in conductance delays.

Consider the network portrayed in Figure 1. If neurons a, b, and c spike at the same time, Time 0, those spikes will be received by neuron x at Times 1, 5, and 9, respectively, and those same spikes

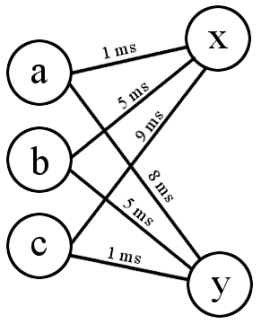

Figure 1: A small neural network with time delays.

will be received by neuron y at Times 8, 5, and 1, respectively. In this case, neither x nor y receive the coincident spikes needed to fire. The difference in arrival times are caused by differences in axonal propagation times. If, instead, neurons a, b, and c spike at Times 8, 4, and 0, respectively, neuron x will receive all three of these spikes at Time 9, potentially allowing the cell to fire. In contrast neuron y would receive the three spikes at Times 16, 9, and 1, respectively, providing it with no coincident spikes to drive an action potential. Alternatively, if neurons a, b, and c fire in the *reverse* order, neuron y will may spike, while neuron x will remain silent. Thus, the effects of spikes from neurons a, b, and c on the firing of neurons x and y is critically dependent on the timing of the spikes.

In larger, more connected, networks, like those found in mammalian brains, a particular stimulus will cause a chain reaction of spikes over time. This group of neurons firing with precise timing is what forms a corresponding PNG. Importantly, PNG patterns can be strengthened with repetition through synaptic mechanisms that exhibit spike timing dependent plasticity (STDP) (E. Izhikevich, 2006). Synapses that exhibit STDP are *strengthened* whenever the post-synaptic neuron fires just *after* it receives evidence of a pre-synaptic spike. Conversely, whenever the post-synaptic neuron fires just *before* it receives evidence of a pre-synaptic spike, then the synapse is *weakened* (Dan, Poo, & et al, 2004). Thus, as a PNG unfolds, STDP strengthens the synapses participating in the PNG's generation and weakens the synapses that were active but did not facilitate the firing of neurons participating in the PNG. Thus, every time a particular PNG unfolds, and hence becomes strengthened via the mechanism of STDP, it becomes easier for that PNG to be reproduced.

As previously defined, a PNG is a reproducible, time-locked, spatiotemporal pattern of spikes. A PNG is reproducible in the sense that, when the neurons participating in a PNG are stimulated in a similar way, the PNG will unfold in a similar way. PNGs become more stable through mechanisms of

synaptic plasticity, such as STDP (Dan et al., 2004; E. Izhikevich, 2006). This increases the likelihood that the PNG will be triggered in similar situations in the future, and it makes the PNG increasingly robust to noise. Some input spikes may be omitted or added without substantially effecting the generation of the PNG. A PNG is time-locked due to the fact that the propagation delays between the participating neurons are fixed by the anatomy of the network. A PNG is spatiotemporal in the sense that it necessarily occurs at many times (polychronous) and involves many neurons. Once stabilized via STDP, subtle variations in spike timing due to noise do not lead to unpredictablely different PNGs, but generate a member of a family of related PNGs (E. Izhikevich, 2006). Also, it is important to note that many PNGs may be simultaneously active in a common neuronal network without interacting, due to the low probability that two arbitrary PNGs will overlap substantially in both the set of neurons involved and their precise spike times.

PNGs also minimize redundancy through the weakening of synapses via STDP, and they are more energy efficient than neural coding schemes that depend on neural firing rate (Levy & Baxter, 1996). It is also interesting to note that the "small world" connectivity structure of the mammalian brain gives rise to stable PNGs much more readily than networks of neurons that are connected uniformly at random (Sporns & Zwi, 2004; Vertes & Duke, 2010).

Since their introduction, PNGs have been utilized extensively in computational neuroscience models of cognitive information processing. The intricate dynamics of PNGs have been used in combination with models of NMDA receptors and neurotransmitter reuptake to produce a promising account of working memory function (Szatmáry & Izhikevich, 2011). PNGs have been incorporated into a formal account of the dopamine system in order to produce a candidate model of neural reinforcement learning that addresses the problem of temporally distal reward (E. Izhikevich, 2007). In addition to their use in computational neuroscience models, empirical evidence for PNGs has been reported, with reproducible, time-locked, spatiotemporal patterns of spikes being observed in cortical slices (Rolston, Wagenaar, & Potter, 2007).

As this review of PNGs shows, the information carried by a PNG in a neural network is critically dependent on the timing of individual action potentials. This contrasts with vector space accounts of mental representation which map vector space dimensions onto the instantaneous firing rates of neurons. The PNG approach highlights the way in which individual spike times can carry information, with spiking rates lacking sufficient spatiotemporal detail to discriminate between different represen-

tational states. It is this shift that changes the way in which similarity is realized in spiking neural networks when viewing PNGs as the foundation of mental representation.

2.2 Dynamical Systems Neuroscience

The dynamics described in the previous section are general to spiking neural networks modeled with action potential propagation delays taken into account. There are many kinds of neurons, however, and there are likewise many ways to model those neurons. Nearly all non-transducing neurons generate action potentials—an all-or-nothing signal dependent on the membrane potential at a spiking neuron's axon hillock. How is this neuron dynamic so predictable even though its components are so complex and variable?

The membrane potential of a neuron is largely driven by changing intracellular concentrations of sodium and potassium ions (A. Hodgkin & Huxley, 1952a). Measurements from large squid neuron electrophysiology experiments were used by Hodgkin and Huxley to create their seminal dynamical systems model of neuron membrane potential over time (A. L. Hodgkin & Huxley, 1952):

$$I = C_M \frac{\mathrm{d}V}{\mathrm{d}t} + \bar{g}_{\mathrm{K}} n^4 (V - V_{\mathrm{K}}) + \bar{g}_{\mathrm{Na}} m^3 h (V - V_{\mathrm{Na}}) + \bar{g}_L (V - V_L), \tag{1}$$

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \alpha_n (1 - n) - \beta_n n, \qquad \alpha_n = \frac{n_\infty V}{\tau_n}, \quad \beta_n = \frac{(1 - n_\infty)}{\tau_n},$$

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \alpha_m (1 - m) - \beta_m m, \qquad \alpha_m = \frac{m_\infty V}{\tau_m}, \quad \beta_m = \frac{(1 - m_\infty)}{\tau_m},$$

$$\frac{\mathrm{d}h}{\mathrm{d}t} = \alpha_h (1 - h) - \beta_h h, \qquad \alpha_h = \frac{h_\infty V}{\tau_h}, \quad \beta_h = \frac{(1 - h_\infty)}{\tau_h}.$$

Where I is the electrical current across the neuron membrane, C_M is the constant membrane capacitance per unit area, V is the variable membrane potential, $\bar{g}_K$ is conductance per unit area of potassium, n is the dimensionless proportion of potassium inside the membrane, V_K is the voltage difference between resting potential and the equalibrium potential of potassium ions, $\bar{g}_{Na}$ is the conductance per unit area of sodium, m is the dimensionless proportion of activating molecules within

sodium channels, h is the dimensionless proportion of inactivating molecules blocking sodium channels, V_{Na} is the voltage difference between resting potential and the equalibrium potential of sodium ions, $\bar{g}_L$ is the conductance per unit area of ions associated with the leakage of current, and V_L is the voltage difference between resting potential and the equilbrium potential of ions associated with the leakage of current. Each m, n, and h have their own α and β rate constants that change with membrane potential, modeling the relationship between m, n and h and voltage. These rate functions can be approximated by considering their limit as t approaches ∞ and adjusting the τ parameter to fit experimental data.

This can describe action potential propagation by combining this model with the relation

$$\frac{a}{2R\theta^2}\frac{d^2V}{dt^2} = I,$$

assuming a constant propagation speed θ, given resistivity, R (A. Hodgkin & Huxley, 1952b). For a particular neuron of interest, numerical simulations can be run by guessing different values of θ. The model diverges to either ∞ or $-\infty$ if the guessed velocity is too fast or too slow. If θ is just right, however, the model will converge to its resting membrane potential. The neighborhood of θ where the membrane potential converges to resting potential serves as a prediction of the velocity of action potential propagation through the nerve fiber.

The change in model behavior with respect to θ is called a *bifurcation*, where changing a parameter qualitatively changes the dynamics of the dynamical system (Poincaré, 1885). Any change to the above listed experimentally measurable values will likewise shift the bifurcation point of the system with respect to θ, and hence shift the range of possible propagation speeds. In general, by analyzing points of bifurcation, we are able to qualitatively characterize the impact of specific parameter changes, and determine ranges of interest for modeling particular dynamics.

The Hodgkin-Huxley electrophysiological model has many known bifurcations which have been studied extensively (E. M. Izhikevich, 2000). Indeed, the system only demonstrates "excitability" when the parameters are such that a bifurcation with respect to membrane potential exists near its stable resting potential (E. M. Izhikevich, 2007). Using these extensive bifurcation analyses, quadratic integrate-and-fire models, and β-reductions in a lambda calculus, Izhikevich made a simple neuron model which preserves the biologically relevant bifurcation dynamics, while reducing the computational complexity of its simulation:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = 0.04v^2 + 5v + 140 - u + I \qquad \frac{\mathrm{d}u}{\mathrm{d}t} = a(bv - u) \tag{2}$$

$$\text{If } v \geq 30 \text{ mV, then } \begin{matrix} v \leftarrow c \\ u \leftarrow u + d \end{matrix} \quad (E.\ Izhikevich, 2003).$$

In this simplified model, v represents the membrane potential of the neuron, u is the recovery variable which tends to drive the membrane potential back to equilibrium. The dimensionless parameters a, b, c, and d shape membrane potential dynamics: a scales the change of the recovery variable, where smaller values slow the recovery of the membrane potential; b determines the sensitivity of the recovery variable to sub-threshold voltage dynamics; c determines the value to which v is set after an action potential (when $v \geq 30$); and d is the amount to increase the rate of recovery after spiking (invoking a refractory period). With different combinations of these 4 parameter settings, the model can produce many significant electrophysiological dynamics. These include regular spiking, instrinsically bursting, chattering, fast spiking, thalamo-cortical, resonator, and low-threshold spiking dynamics (Figure 2).

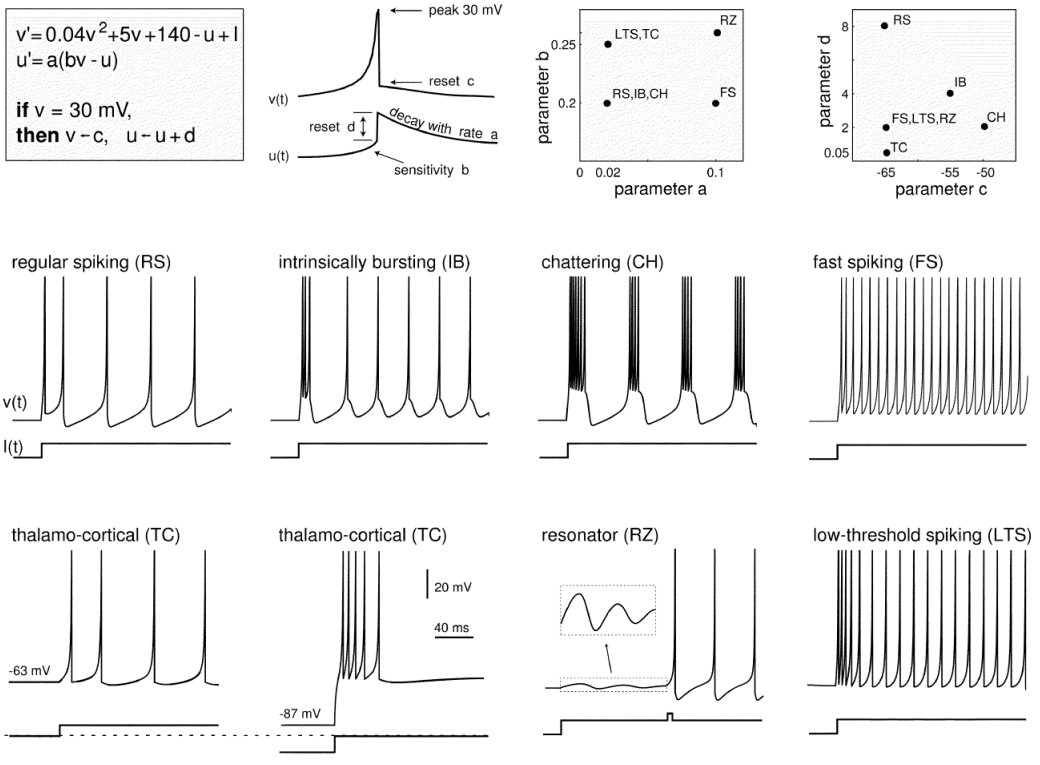

Figure 2: The relationships between particular parameter settings and model dynamics in response to 10 mV dc stimulation. RS, IB,TC, and CH are excitatory, while FS and LTS are inhibitory.
(E. Izhikevich, 2003)

It is important to recognize, however, that Hodgkin-Huxley type neuron models are only predictive of a neuron's electrophysiology around the axon hillock—further details of the neuron model, including but not limited to action potential propagation dynamics, synaptic dynamics, and dendritic integration dynamics. Since the model is based on electrophysiological dynamics, and not explicitly implementing the theoretical mechanisms that give rise to those dynamics, its use depends on the prudence of the modeler to interface it with other biophysically meaningful models. Polychronization across regular spiking neurons in a network depends critically on the modeling of action potential propagation speeds, and the robustness of these dynamics depend on modeling synaptic plasticity.

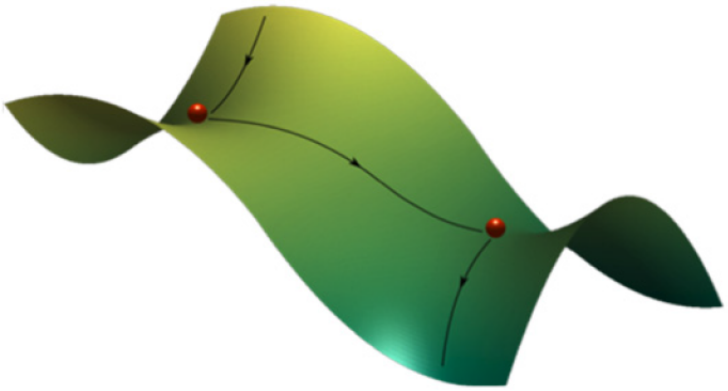

Figure 3: A heteroclinic chain of saddle nodes. This is the simplest possible heteroclinic chain. (M. I. Rabinovich, Afraimovich, Bick, & Varona, 2012)

In dynamical systems theory, a *limit cycle* is a closed (periodic) trajectory in phase space which can be approached from another trajectory, but continues to cycle as time progresses to ∞. A limit cycle can be: *stable*, where trajectories from all sides approach the limit cycle as time goes to $+\infty$ (like a stable attractor); *unstable*, where trajectories from all sides approach the limit cycle as time goes to $-\infty$ (trajectories are repelled going to $+\infty$); and *semi-stable*, where a trajectory from at least one side approaches the limit cycle when time approaches $+\infty$, but a trajectory from another side approaches the limit cycle when time approaches $-\infty$ (the limit cycle can only be approached by trajectories from some directions).

Considering the dynamics of larger networks, the bifurcations exhibited by individual neural dynamics shape trajectories of the whole system. The limit cycles of spiking neurons exhibit bifurcation with respect to variation of the tonic input current. A *saddle node* bifurcation is one where variation

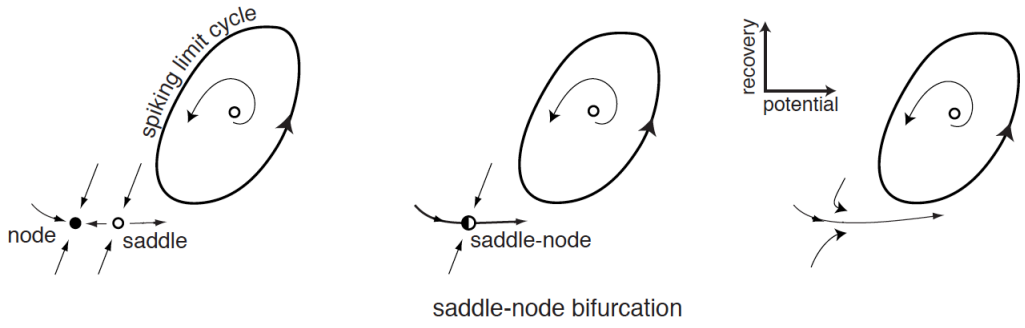

Figure 4: Phase portraits of Equation 2.2, showing a saddle node bifurcation with respect to tonic input current. On the left, many values of membrane potential converge to a fixed point resting potential. On the right, all values approach a stable limit cycle that results in a frequent spiking dynamic.
(E. M. Izhikevich, 2007)

of a parameter causes two fixed points to collide, one stable and one unstable, destroying each other. Moving the parameter in one direction shifts the system into a regime with two fixed points, while moving the parameter in the opposite direction shifts the system into a regime with no fixed points. When current is too low, a neuron's dynamic tends to a resting potential as time approaches ∞, a stable fixed point (where the dynamics surrounding a spiking event occur as a trajectory approaches a corresponding unstable fixed point). When tonic current is sufficiently increased, a neuron consistently builds membrane potential, exhibits a spike, then resets back to resting potential as it builds up more membrane potential, repeating the process. This oscillation between spiking and resetting is a limit cycle to be repeated indefinitely as time approaches ∞. The spiking dynamic of an action potential corresponds to a saddle node bifurcation, because the membrane potential rapidly accelerates toward a threshold value as a result of increased input current, after which it is repelled back to resting potential. This limit cycle will be maintained as long as the input current is maintained.

A *heteroclinic chain* is a trajectory of states that pass through multiple metastable states, like the ball portrayed in Figure 3, which passes through two saddle nodes before being propelled away from the final saddle. In some neural networks, if the action potential from one neuron causes an action potential in another, then the combined dynamic may constitute a heteroclinic chain, in that the current stimulating the first neuron causes a saddle bifurcation, leading to an increase in current in the second neuron, which may cause a second saddle bifurcation. If the current stimulating the first neuron decreases, then so does the heteroclinic dynamic rippling through the coupled neurons. The robustness of that heteroclinic chain is then related to the degree to which conditions are such that

the system may exhibit the necessary saddle node bifurcations. If there is a neighborhood in which the necessary dynamics are guaranteed to propogate the heteroclinic chain, then these heteroclinic trajectories form a *stable heteroclinic flow* through a *stable heteroclinic channel*, portrayed in Figure 5 (M. Rabinovich et al., 2001; M. I. Rabinovich et al., 2012).

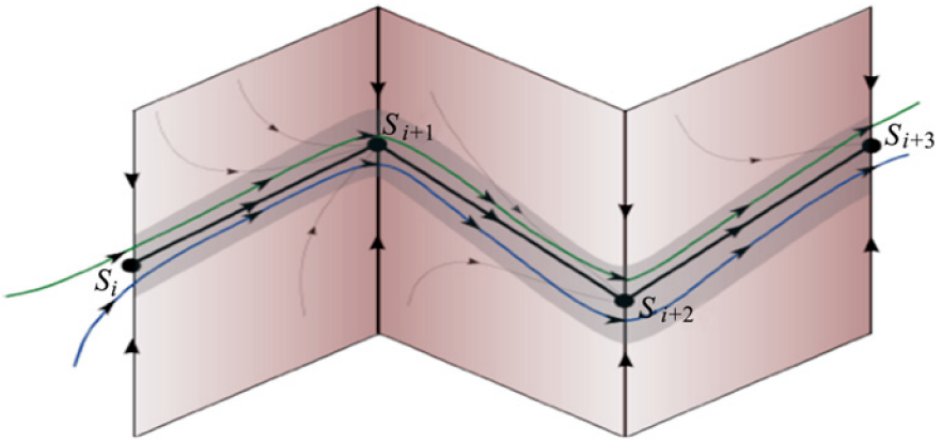

Figure 5: In this figure, heteroclinic chains flow through a larger heteroclinic channel, plotted in phase space. The points in this channel describe the robust metastable states through which dynamics are stereotyped.

(M. I. Rabinovich et al., 2012).

The stabilization of attractive neighborhoods approaching saddle nodes constituting a heteroclinic channel result in dynamics that are reproducible, and may be robust to noise (depending on the size of the neighborhood with respect to noise). Accounts that identify the stability of PNGs as being reinforced through synaptic STDP processes (E. Izhikevich, 2006) may then be seen as the formation of a stable heteroclinic channel, where a particular occurrence of a PNG constitutes a heteroclinic flow of saddle node bifurcations of the membrane potential trajectories of neurons participating in the chain. This characterization forms a useful foundation for characterizing neuronal network dynamics

3 Emergent Phenomena

I wish to propose a severe restriction of the term 'emergence,' calculated to remove this disquieting subjective sense of the mysterious - William Marias Malisoff, 1939

Cognitive scientists often describe cognition as *emerging* from neural processing (McClelland & et al., 2010). How this emergence occurs is uncertain. What does it mean for something to be *emergent*? When has something *emerged*? O'Conner and Wong describe *emergent properties* as those which 'arise' out of more fundamental entities and yet are 'novel' or 'irreducible' with respect to them (O'Connor & Wong, 2012). Here, 'arise' is used because of what it suggests–that it is by the functioning of these fundamental entities that these properties are birthed. Yet, what has arisen is believed to be new or distinct–'novel' or 'irreducible.' We intuitively claim emergence when we ask how a theory at one level of analysis gives rise to a theory at an adjacent level of analysis, i.e. we attempt an intertheoretic reduction.

In this section, I will show that the observation of emergence is dependent on the particular historico-theoretical context of the observer, leading to the reflection of this notion in the contexts of meaningful emergent dynamics (Section 4.1). In this context, what appears to be emergent are cases of intertheoretic reduction that are complex or difficult to conceptualize. I will clarify this by showing a formalization of the relationship between levels of analysis grounded in dynamical systems, using what I will call "$\phi\rho\psi$" relations. $\phi\rho\psi$ relations will be demonstrated in the domain of gliders in the Game of Life, as well as for polychronous neuronal groups in spiking neural networks. I will conclude with the illustration of the use of these $\phi\rho\psi$ relations to aid in the grounding of polychronous neuronal groups as a unit of representation for the brain, and apply this grounding to describe the development of language awareness.

3.1 Emergence in Philosophy

Early approaches regarding emergence arose out of discussions of how to ground biology and chemistry in physics. This discussion was in part forced by their parallel advancement, each leveraging wildly different vocabularies. This discussion yielded mechanists, who emphasized how all things more complex are reducible to physics, and vitalists, who posited an "entelechy"–an innate force that guided actions throughout development (Driesch, 1909, 1914).

Of the mechanists, there were three notably influential writers: Broad, Mill, and Alexander (Broad, 1925; Mill, 1843; S. Alexander, 1920) (O'Connor & Wong, 2012). Broad and Mill wrote of a form of ontological emergence, or supervenience emergentism. In this view, emergence implies the coming into being of a new, fundamental property, with new causal powers. In contrast, Alexander wrote of an epistemological emergence, the emergence we discuss throughout this paper. Epistemological emergence only commits to the higher level appearence of novelty; epistemologically emergent properties do not exist beyond our reference to them in human knowledge.

One example of emergence is perceptual experience arising from brain function. In 1714 Leibniz posited a mechanistic thought experiment to ponder notions of philosophy of mind:

> Moreover, it must be confessed that perception and that which depends upon it are inexplicable on mechanical grounds, that is to say, by means of figures and motions. And supposing there were a machine, so constructed as to think, feel, and have perception, it might be conceived as increased in size, while keeping the same proportions, so that one might go into it as into a mill. That being so, we should, on examining its interior, find only parts which work one upon another, and never anything by which to explain a perception. Thus it is in a simple substance, and not in a compound or in a machine, that perception must be sought for (Leibniz, 1714 and 1898).

Leibniz paints the portrait of a huge, conscious mill to imagine how it might appear. He concludes that since he cannot see "anything by which to explain a perception," that there must be a "simple substance not in a compound or in a machine that must be sought for." This has been termed Leibniz's gap. Leibniz is right to doubt the dynamics of the mill. The mill that he imagines has simple mechanisms, built of wood or metal, with relatively stable dynamics. Biological function is dependent on the unique properties of elements and molecules, which have been structured by evolution to perform biological metabolic processes. Yet, from the imagined mechanisms of the mill, conscious perception is to rise. For the mechanisms of a mill to have the dynamics of the mechanism of a neuron, the components of the mill need initially be assembled to "simulate" the dynamics of a neuron. For Leibniz to successfully experiment with thought, his intuition would need to include structures greatly overcomplicated by the nature of the mill's fundamental components. For us to successfully experiment with thought, and overcome Leibniz's gap, our intuition must be in terms of the components from whose dynamics perception is natural.

There are two key distinctions discussed regarding emergence: irreducibility, and unpredictability. To say an account of an emergent property is irreducible is to say that it cannot be reduced to a fundamental physical theory. Jerry Fodor stresses this point by supposing an 'immortal economist' who believes he can explain economic theories from his understanding of physics–something Fodor argues as fruitless and overcomplicating, hence the economist's immortality (Fodor, 1974; Nagel, 1961). Taken weakly, a claim of irreducibility is an issue of the conceptual limitations of a given underlying theory. Taken strongly, a claim of irreducibility suggests a sort of magic that mysteriously disconnects a theoretical account of an emergent property from an underlying physical explanation.

To address this distinction, Mark Bedau distinguishes between *nominal*, *weak* and *strong* emergence. *Nominal* emergence refers to a property that appears to be inexplicable, or is surprising to a human observer (Ronald, Sipper, & Capcarrere, 1999). This puts no constraints on what is emergent other than the human capacity to declare something as such. *Weak* emergence refers to an emergent property that can only be explained after or by simulation, and hence is either unpredictable or difficult to predict (Bedau, 1997). This characterization of emergence likely includes most serious examples of emergent properties. *Strong* emergence refers to an emergent property that, genuinely, has its own lawlike functions that are wholly seperated from underlying theories, and therefore cannot be simulated by modelling those underlying theories. This characterization enables *downward causation*, since the composition of the substance that exhibits emergent properties cannot be modeled to simulate the emergent properties, the emergent properties must be taken to *cause* dynamics of the underlying system–akin to the strong irreducibility claim (Kim, 1999). However, nominal emergence is subjective (Nagel, 1961)–what we claim is emergent is dependent on the knowledge of the observer. Thus, we take it as prudent to recognize that properties which appear emergent are not necessarily strongly emergent. For this reason, the appearence of downward causation can be thought of as *downward confinement*, where the explanation of the overlying system helps us constrain its explanation in the underlying system (Atmanspacher, 2011). William Wimsatt, a *New Mechanist*, pragmatically emphasizes the importance of understanding the way that features of a system contribute to it appearing weakly emergent or *non-aggregative* (Wimsatt, 1997). This approach to understanding the ontology of emergent systems is argued to help safeguard against "lazy" reductionism in cases where the appearance of complexity and the sensitivity to details may otherwise *demotivate* attempts to clarify details of that reduction (Wimsatt, 2006). This argument, combined with downward confinement, implies

that the ontology of a system may be seen in terms of the properties that lead to dynamics appearing emergent, while those epistemological properties can be seen as constraining related epistemological claims.

3.2 Emergence, Theory-ladenness, and Historico-theoretical Context

Our perception is *theory-laden*; what we observe is in terms of what we have experienced (Kuhn, 1962; Estany, 2001). Our theory-laden perception has been revealed in many contexts: our ability to judge similarity between perceived colors is dependent on our culture and language (Whorf, 1956a, 1956b; Gilbert, Regier, & Ivry, 2006), our ability to identify spoken phonemes is dependent on our visual experience (McGurk & MacDonald, 1976)–even our ability to perceive the steepness of slopes has been argued to be dependent on how encumbered we are (Proffitt, 2006). Further, our perception is categorical; our knowledge about the world structures what it is possible to perceive (Oppenheiml & Putnam, 1958; Harnad, 1990). Our knowledge of the world is ever changing. The quality of our perception seems inextricably bound to the distinct historico-theorical context it is experienced in.

It seems that particular historico-theoretical contexts favor some perceptions, and obscure others. If we view theories of emergent properties in terms of the accessibility of their intertheoretic reduction, then what we claim as a mysteriously emergent phenomenon is dependent on having a particular historico-theoretical context that either favors or obscures the perception of the theoretical domains between which the intertheoretic reduction is to be established. This notion seems to imply that some historico-theoretical contexts lead to some phenomena seeming more or less intuitively reducible. Let us consider phenomena that are described as emergent.

In 1932, fluid dynamicist Horace Lamb is famously reputed to have said:

> "I am an old man now, and when I die and go to Heaven there are two matters on which I hope for enlightenment. One is quantum electrodynamics and the other is turbulent motion of fluids. And about the former I am really rather optimistic." (Tabor, 1989)

In this quote, Lamb reveals his belief that the emergence of turbulence is so strong that even a god could not explain it. While I will not make Lamb's commitment, it is clear that in our historico-theoretical context turbulence in fluids is *weakly emergent*. Considering our characterization of emergence as intertheoretic reduction, turbulence in fluids is *very* difficult to reduce to our current theory of fluid dynamics.

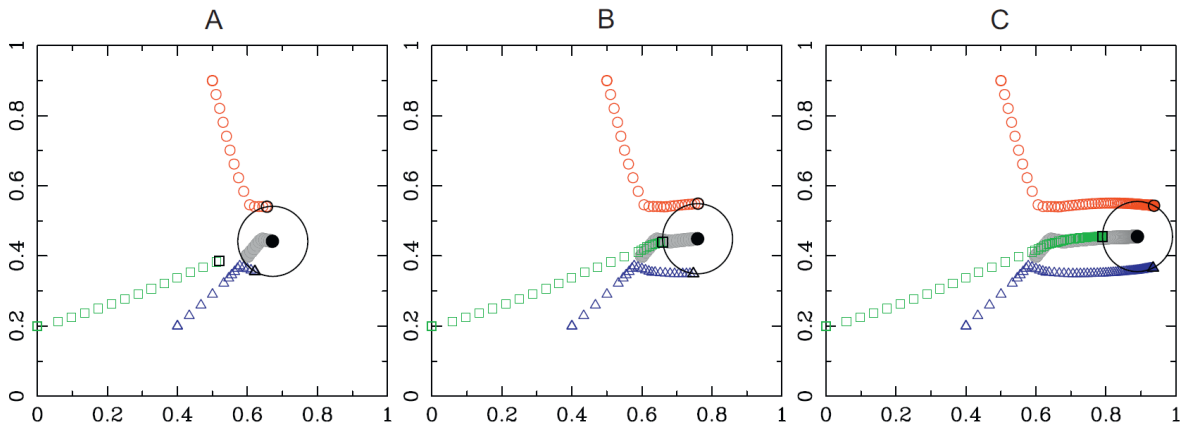

Figure 6: Wolf-pack of three wolves, initially randomly distributed, hunting a reactive prey. (A) Approaching phase and start of the pursuit. (B) Pursuit and encircling phases. (C) Pursuit, encircling and capture. The solid black circle is the prey, and the unfilled shapes are the wolves. Once they get a "safe distance" away from the prey (the black ring around the prey), they move to get equidistant from each other wolf who is a "safe distance"" away.
(Muro, Escobedo, Spector, & Coppinger, 2011)

Yet, the appearence of complexity does not ensure weak emergence. *Canis lupus*, wolves, are credited with having some of the most complex social behaviors of Carnivores (Macdonald, Creel, & Mills, 2004). One important motivation for this ascription of social complexity is their intricate pack hunting behaviors, where wolves track prey, carry out pursuit, and encircle that prey until it stops moving. In 2003, Peterson and Ciucci surveyed well-known wolf biologists regarding their opinions regarding the intentionality of wolves hunting in a pack, where 15 out of 17 of those surveyed agreed that wolves used cooperative hunting strategies (Peterson & Ciucci, 2003). Indeed, it is difficult to reduce pack hunting behavior to our theory of individual wolf cognition. However, recently it has been shown by simulation that for a wolf pack to have these hunting behaviors, it is sufficient for each wolf to 1) move toward the prey until a minimum safe distance to the prey is reached, and then 2) move away from the other wolves that are close to the safe distance to the prey (see Figure 6) (Muro et al., 2011). This explanation of wolf-pack hunting behaviors does not depend on assumptions of high levels of wolf intelligence, and seems to match our expectations for the behavior of a wolf. Without knowledge of a simple solution, the perception of pack hunting behavior as cooperative and complex is favored as intuitive. As our historico-theoretical context changes, the difficulty of understanding the intertheoretic reduction changes.

Viewing accounts of the emergence of a property in terms of their intertheoretic reduction suggests that our understanding of emergence would benefit from a formal account of intertheorhetic reduction.

3.3 Emergence in Dynamical Systems

We can describe the attempt to explain how a property emerges as *intertheorhetic reduction*. An intertheoretic reduction is the illustration of how one theory near perfectly predicts another. For a given emergent property, there are at least two levels of description: the underlying system that does not recognize the existence of the emergent property (e.g., physics), and an overlying system that does recognize the existence of the emergent property (e.g., chemistry). These underlying and overlying systems are often characterized as *dynamical systems*, a mathematical concept where a function determines the progression of time through an abstract space. This characterization has fueled recent thinking in philosophy of on the emergence of mental states from neural dynamics. Atmanspacher, Butterfield, Rosaler, and Yoshimi each describe the relationship between explanatory levels through a form of *coarse graining*–a term from Statistical Mechanics that implies a way of dividing the underlying space and mapping it onto the overlying space (for Atmanspacher, Butterfield, and Yoshimi: contexts, meshing, bridge maps, and supervenience functions, respectively) (Atmanspacher, 2011; Butterfield, 2012; Yoshimi, 2012b). A coarse graining is a way of describing groups of states in the underlying system that, for the purposes of explanation and prediction, are equivalent. For each of Atmanspacher, Butterfield, and Yoshimi, this description is essential for discussing how neural dynamics give rise to, and correspond with, mental states. Here, I introduce the $\phi\rho\psi$ framework in order to clearly and flexibly discuss intertheoretic reductions.

In this framework, we define three functions, ϕ, ρ, and ψ. ϕ is the underlying dynamical system. Its space may be conceived as fundamental, and the rule that determines its progression over time can be thought of as the implementation of physical laws, or a model of our physical theories. In this space, we may identify ρ, a function that serves as a detector for the existence of some property (or set of properties) in the space that ϕ operates on. ρ becomes identical to the *definition* of the property in ϕ and constructs an overlying space ψ. ψ is the overlying dynamical system. Its space is, by the definition of ρ, descriptive of ϕ. The rule that determines its progression over time is relative to ρ and instantiates our theories of how ρ functions. By identifying a property of ϕ, we implicitly define ρ and create an overlying system ψ. As a definition, we can think of ρ as implying equivalence classes for the overlying system ψ, providing a method of coarse graining the underlying system ϕ.

3.4 Gliders in the Game of Life

A clear illustration of this framework for intertheoretic reduction can be shown with John Conway's Game of Life (Gardner, 1970). The Game of Life is a cellular automata dynamical system that has many spontaneously emerging structures. It simulates a society of living organisms; each automata, or cell, has position implied by its location on a grid of arbitrary size, and lives or dies as a function of population density of adjacent cells. If a dead cell position has exactly three live neighbors, it becomes alive as if by reproduction. If a living cell has two or three live neighbors, then it stays alive. In all other conditions, a cell dies or remains dead. The Game of Life is considered an example of how complex structure emerges from simple rules. Indeed, it has been shown that the living cells in an arbitrary grid of the Game of Life exhibit self-organized criticality (Bak, Chen, & Creutz, 1989).

The Game of Life has emergent living structures that can be wholly explained by the underlying rules of the system, but seem to call for further explanation. To make this clear, let us consider a glider. A 'glider' is a pattern of living and dead cells that, if uninterrupted, self-replicates every other iteration in a position slightly shifted from the previous [Figure 7]. This shifting gives rise to its name, glider, as it results in the appearance of a shape of five living cells gliding across the grid [Figure 8].

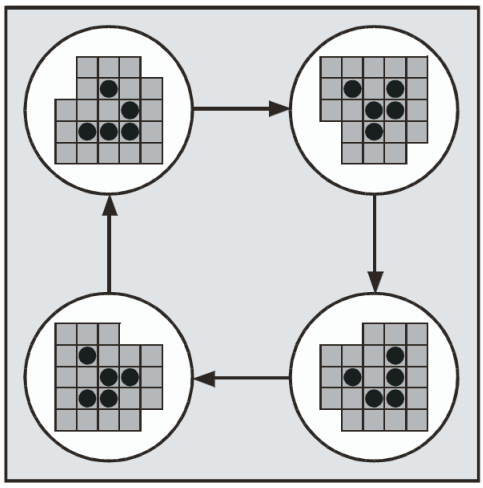

Figure 7: Life of a glider. Every nth iteration is a transposition of the (n-2)th iteration. (Beer, 2004)

The glider has a long history of being studied in the context of emergence (Bak et al., 1989; Standish, 2001; Bedau, 1997), including a consideration of the 'cognitive domain' of its interactions (Beer, 2004). Here, we are principally interested in the glider as an example a spontaneously forming emergent structure that can and has been studied, whose domain of existence is quantifiable and

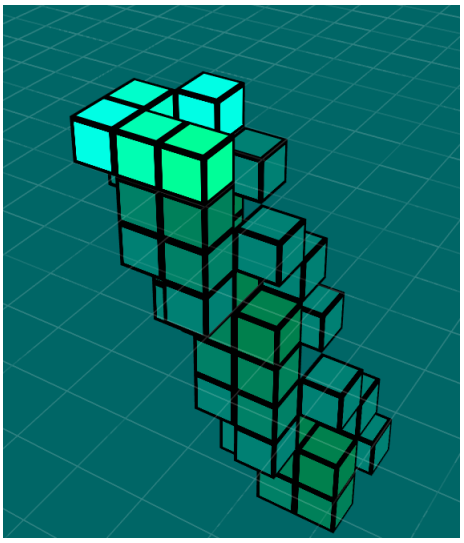

Figure 8: Self-Replication results in the appearance of movement across the grid. Above, the shaded boxes represent living cells in past iterations. (Xerol, 2009)

is clearly relatable to the original system. Theories of cognition describe emergent mechanisms of information flow in the brain, as theories of glider movement describe emergent mechanisms of cellular automata updating in the Game of Life. We expect the insights gained by relating glider study to the Game of Life to inform relating cognitive science to neural function.

3.5 $\phi\rho\psi$ Analysis of Gliders

$\phi(s_0) = s_1$	Phi, a dynamical system operating on the source space which represents the causal forces of the Game of Life that maps a grid state, s_0, onto its subsequent grid state, s_1.
$\rho(s_0) = e_0$	Rho, a definition of a glider that maps a grid state, s_0, onto glider space, e_0.
$\psi(e_0) = e_1$	Psi, a dynamical system operating on glider space that maps a state, e_0, onto its expected subsequent state, e_1.

Figure 9: Emergent structures are embedded in the source space, the original space with independent causal forces. ϕ represents the application of theories on the source space s_0, instantiated by a model. This application yields a future source space s_1. Our theories of relevant emergent structures from the source space imply the existence of ρ, which maps from the source space s_0 onto a new glider space, with state e_0. ψ represents the application of our emergent structure theories on e_0. This application yields a future state of glider space, e_1. This characterization is guaranteed by the identification of an emergent structure, and implies relationships between all theories of emergent structure.

Let us consider the Game of Life without full knowledge of its function. In this way, we treat the Game of Life the way we treat our world: the world is, and functions via the causal forces of the world. Here, we refer to causal forces as being whatever force happens to change the grid. As time progresses on the grid, the change we see can be described by the grid's causal forces. We can try as

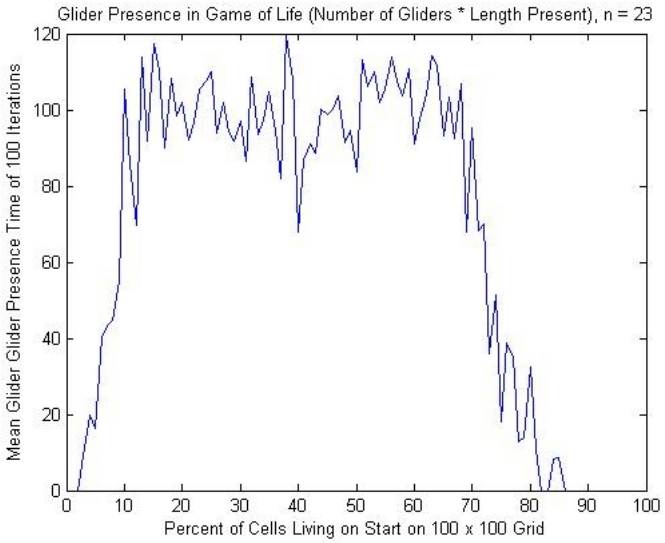

Figure 10: Gliders are a natural emergent structure in the Game of Life. Each glider is born, exists for some time, then dies. Above, we sum up the lifetimes of all gliders that occurred over 100 iterations. What is plotted is the mean of those sums across 23 different randomized grids. This is identical to, for each iteration, summing the number of Gliders detected by ρ. Thus, the plot is proportional to the probability that a Glider is present on any given iteration, where points over 100 imply more than one Glider.

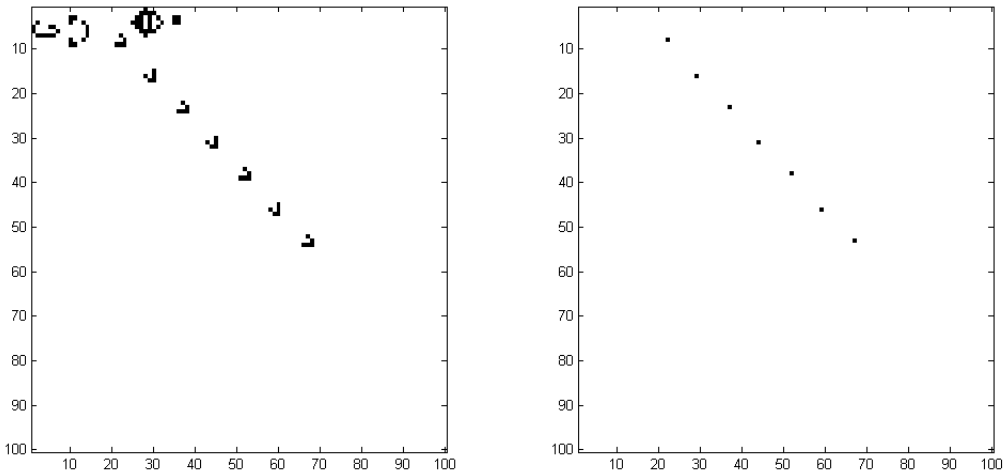

Figure 11: On the left, we see a Game of Life grid containing a Gosper Glider Gun 200 iterations in. It has produced 7 Gliders. On the right, we see glider space. Each dot represents the presence of a Glider centered on that location.

we may to describe these causal forces: let ϕ be our simplest and finest description of these causal forces, where ϕ represents a function which maps any state of the Game of Life onto its succeeding state. ϕ implements the rules of the Game of Life as we have defined them, operating on grid states in the system, s. A grid state, s, can be represented as a square matrix of arbitrary size with integer

modulo-2 elements, where 1, or 0, implies the existence, or non-existence, of a living cell at a position implied by that element's position in the matrix.

In this world, we observe the Game of Life. As we do, tendencies emerge, we notice what we call gliders and we seek to understand the causal structure of their dynamics. Gliders occur spontaneously from random initial grids [Figure 10]. As long as the population density of the initial grid is between 1000 and 6500 live cells per 10000 cells, the sum of all the lifetimes of all gliders spontaneously occurring nears 99.5 iterations if that grid is iterated 100 times. Thus, gliders are an active and near constant presence in even random grids. It seems gliders are a natural dynamic of cell life in the Game of Life, yet their behavior, while fully explainable by ϕ, is not contained in the description ϕ. By identifying the existence of gliders, we produce a new space that can be used to more efficiently explain their dynamics.

The space the gliders exist in is necessarily contained by the grid of the Game of Life; and its elements are fully dependent on the elements of the grid. In this way, the glider space is embedded within the Game of Life [Figure 11]. Much like how ϕ described the mapping of any state of the Game of Life onto its succeeding state, we can define ψ such that it describes the mapping of any state of glider space onto its succeeding state. We can define a relationship as ρ; ρ is a mapping from the grid state onto glider space. ρ thus becomes representative of how we define gliders, and its application instantiates a glider detector. In Figure 11, we have created and applied ρ, and plotted its results with the Game of Life grid on the left and glider space plotted on the right. On the left, the grid contains a 'Gosper Glider Gun', a small structure that produces a Glider every 30 iterations. After 200 iterations, it has produced seven gliders; their grid structure is visible on the left. With the application of ρ, we see the glider space formed on the right. As gliders move, we see their corresponding points of existence move. By identifying the existence of gliders, we imply the glider space and we begin to perceive it. On viewing, it becomes clear that once a glider exists and is in motion, it stays in motion until it no longer exists. Gliders always move in the same direction, but the dimension they traverse over varies every other iteration, resulting in diagonal movement. Since the existence of gliders in future iterations is dependent on information not available in glider space, complete descriptions of the dynamics of ψ cannot be wholly independent of the underlying system. Given our ϕ and ρ, ψ in this case is at least a discrete markov chain and at best a dynamical system, where ψ predicts probablistic outcomes from e, the state space that ψ operates on. We can think of this as corresponding to states of phenomenal

existence (Norris, 1998).

Let us summarize our $\phi\rho\psi$ description of gliders in the Game of Life [Figure 9]. In the Game of Life, we have a lattice of arbitrary size. Time progresses as the rules of the Game of Life, ϕ, are applied to this grid; $\phi(s_0) = s_1$. We define a glider as any pattern of on/off cells that can be flipped and/or rotated to be identical to the pattern in Figure 7. From this definition, a glider detector can be formed, ρ, that observes the grid, s_0, and defines a state e_0 that corresponds to locations where the existence of a glider has been detected; $\rho(s_0) = e_0$. This new space, e_0, can be described using a new function, ψ, that corresponds to our understanding of the rules by which time progresses in e_0; $\psi(e_0) = e_1$. The glider appears emergent as ψ seems to have simple rules that are not contained in ϕ. Furthermore, it may seem overcomplicating–and inefficient–to simply define $\psi(e_0) = \rho(\phi(s_0))$ to explain the progression of gliders, even though it may be the only function that can precisely predict e_1.

3.6 Relationships between Mappings

ϕ, ψ, and ρ, given the grid, represent models instantiating our theories of how these systems function. Since the glider space is embedded within the Game of Life grid, we assume that there is a relationship between the mappings that describe them. For Figure 11, we did not need to have an explicit definition of the contents of ψ because we had the state s_0, ϕ, and ρ. The behavior of gliders became implicitly defined by knowing the definition of the glider, knowing how the Game of Life works, and knowing the present state of the grid.

The Game of Life and the glider are interesting because we have full knowledge of the system, and the definition of a glider is well known and unwavering. The structure of ρ given ϕ is dependent on our understanding of what we think causes the phenomenon. As a result, we can use the relationship between the functions to assess their consistency.

$$
\begin{array}{ccc}
s_0 & \rho \longmapsto & e_0 \\
\phi \downarrow & & \downarrow \psi \\
s_1 & & e_1 \\
\vdots & & \vdots \\
\phi^n \downarrow & & \downarrow \psi^n \\
s_n & & e_n
\end{array}
$$

Above, we start with the state s_0 and use ρ to determine the state e_0. Then, ϕ operates on s_0 n times, and ψ operates on e_0 n times, yielding states s_n and e_n. If there exists a norm $\|\cdot\|$, we can consider error for a state at a point in time, ϵ:

$$\epsilon = \|\rho(s_n) \cdot \psi^n(e_0)\|.$$

As $\epsilon \to 0$, $\rho(s_n) \to \psi^n(e_0)$. If we assume a perfect ψ, then ϵ implies the accuracy of ρ. Likewise, if we assume a perfect ρ, ϵ implies the accuracy of ψ. Considering them both imperfect, ϵ represents the quality of the description of the phenomena. As the number of ρs used to describe the system increases, the potential for error increases. This becomes intuitive if we think of ρ as corresponding to a level of description. If our prediction is dependent on many levels of description, such as predicting the stock market from physical theories, then sensitivity to conditions at each scale can exacerbate errors in the final prediction.

ψ is constrained and produced with respect to ϕ. This truth enables us to relate beliefs across scales. ψ is evident in the dynamics of ϕ given s_0. Thus, any property of the dynamics of ϕ that makes contact with ρ will be evident in ψ. A property of the system will thus be present in a scale if ρ measures some aspect of that property. If we have imperfect knowledge of ϕ, ρ, and the space ϕ operates on, then we can use emprical evidence to guide our understanding of these functions. We can verify and constrain these functions by leveraging their relationships. Here, I conjecture that this process can be repeated for every space and emergent structure that occurs in nature.

A property can be implicitly defined by what is suggested to exist either perceptually or theoretically. We do not need to know the properties of ϕ or ρ for ψ to exist; we can perceive gliders

without knowing the rules of the Game of Life. Furthermore, we can theoretically be aware of gliders before we perceive them or know their shape.

3.7 $\phi\rho\psi$-relations and Pluralism

There are many ways to describe the world. Because there are many descriptions of the world, there are many explanations of the world. This fact yields a plurality of scientific domains, each with distinctly seperate assumptions for what is fundamental. Many of these explanations may overlap, in the sense that many explanations may exist for the purpose of explaining similar phenomena. This overlap builds conflict with a philosophy of fundamentalism, as it has in cognitive science (Cartwright, 1999). Fundamentalism is a philosophy of science that implies the existence of a single, victorious theory that explains all of the phenomena deemed relevant by the field. Contrasting with fundamentalism is pluralism, a philosophy of science that involves embracing the plurality of explanations, identifying their usefulness with regard to explaining particular phenomena (Kellert, Longino, & Waters, 2006). This approach has been suggested for cognitive science (Dale, 2008; Abney et al., 2014).

Pluralism, when recognized in terms of $\phi\rho\psi$-relations, becomes analogous to considering different levels of description ϕ which enable unique definitions ρ that lend themselves to creating scientific domains ψ that are particularly efficient in a subdomain. Because having different definitions will yield different systems ψ, as a field we are left to consider how each description can inform each other description. In order to compare, we must consider the assumptions that imply the underlying system ϕ. We do not compare biology with quantum physics, because quantum physics does not consider the existence of stable molecules as a level of description. While there are plural descriptions of cognitive theories, most find their instantiation as fundamental: cognitive neuroscience. Part of the reason why there are so many cognitive theories is a result of a disconnect between neuroscience and the phenomena that we seek to explain. This disconnect yields a gap that we endeavor to close, where our endeavors may yield a plurality of theories.

With regard to fundamentalism, $\phi\rho\psi$-relations themselves only imply the existence of the underlying space, upon which a plurality of scientific domains may be theorized. Thus, a fundamentalist must justify their priority for a particular scientific domain as favored over others, perhaps by the efficiency with which that scientific domain may predict phenomena, or by its ease of explanation.

3.8 Emergence in Neural Systems

In Section 2.2, it was described how the Izhikevich model (Equation 2.2) preserved most of the important dynamics present in the Hodgkin -Huxley model (Equation 1). The most fundamental of these dynamics is the emergence of a neuron's action potential. So different are the properties post-spike that, in order to reduce computational complexity, the Izhikevich model identifies the existence of a spike (using the threshold $v \geq 30mV$) as a critical point to apply new context-dependent dynamics ($v \leftarrow c$ and $u \leftarrow u + d$). Since the existence of an action potential in a neural system is so contextually distinct from when an action potential does not exist, this motivates the consideration of an overlying space which describes the precise timing of action potential initiation in the system. The underlying space, ϕ, would be the ensemble of models describing the neural system (including dendrite, axon, and synapse models for the system). Modeling the membrane potential of neurons using the Izhikevich model in Equation 2.2 creates a natural ρ mapping definition which simply identifies the timing of the occurrence of $v \geq 30mV$, since that corresponds with the depolarization dynamic and the propagation of an action potential in the model. The overlying space ψ would then correspond to a description of the neural system in terms of the existence or non-existence of spikes at particular times, offloading the impact of the spike to the model's interface with other components of the system.

When modeling variability of action potential propagation delay along the axon, the timing of spike arrival is not identical to when the action potential was initiated. The pattern of spike arrivals for a neuron are, among other factors, dependent on the length and width of the axon. The causally efficacious heteroclinic chains through the model will then be constrained by the connectivity and location of neurons in the system. The heteroclinic chains which are supported by the structure of the network are called *structural polychronous groups* or *supported polychronous groups* (Martinez & Paugam-Moisy, 2009), and these can be shown to emerge from the network from stimulation (E. Izhikevich, 2006).

Martinez and Paugam-Moisy characterized PNGs in terms of the precise stimuli that evoked them, imitating the $3-tuple$ probing method originally used to demonstrate their emergence (E. Izhikevich, 2006). The conditions that activate a particular PNG can be formally described. A σ-triggered polychronous group refers to the spike-time pattern generated by chain reaction, where σ is the number of spikes required to trigger the PNG (Martinez & Paugam-Moisy, 2009).

Definition A σ-triggered polychronous group refers to the set of neurons that can be activated by a chain reaction whenever trigger neurons $N_k (1 \leq k \leq \sigma)$ fire according to the timing pattern

$t_k (1 \leq k \leq \sigma)$, where σ is the number of spikes required to trigger the PNG (Martinez & Paugam-Moisy, 2009).

With a synapse model that takes into account spike-time dependent plasticity, the impact of neurotransmitter release on the membrane potential of the receiving neuron is either *increased* or *decreased* according to whether a spike arrives just *before* or *after* an action potential of the post-synaptic neuron (Figure 12).

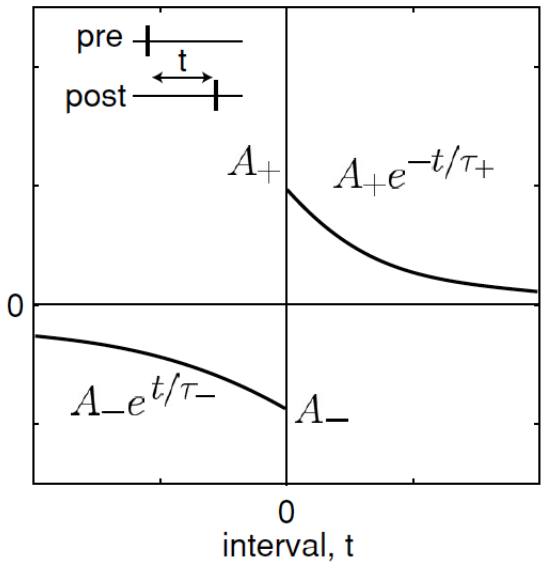

Figure 12: An STDP model where synaptic strength changes based on the relative time of pre-synaptic and post-synaptic action potentials.
(E. Izhikevich, 2006)

The bifurcation of synaptic dynamics with respect to arrival time causes an efficacious synapse participating in the propagation of the heteroclinic chain to strengthen, and those synapses simultaneously active—but not contributing to the propagation of the chain—to be weakened. Changing the synapse strengths also changes what heteroclinic chains are possible. The polychronous groups that are possible as a result of these changes are called *adapted polychronous groups*.

Stability of heteroclinic chains depends on there being a vicinity where all trajectories in a neighborhood of metastable states remain (M. I. Rabinovich et al., 2012). By synapses causing larger post-synaptic potentials, they are expanding the temporal neighborhood wherein coincident spikes must arrive to continue the chain, and they are expanding the likelihood of continuing the chain by requiring of fewer coincident spikes—essentially relaxing the conditions of propagation.

As a result of the shifting stability dynamics of heteroclinic chains from synaptic STDP, and the dependence of those dynamics on action potential propagation delays throughout the network, the polychronous groups that overlap with the activated heteroclinic chains which propagate will see increased stability of all heteroclinic channels which prescribe dynamics. In this way, stable PNGs emerge from the dynamics of spike generation, spike propagation, and synaptic modification. The implications of the formation of these channeled dynamics on the flow of neural information are considered in Section 4.1.

4 PNGs as Mental Representation

An organism can be described in terms of how it reacts in particular contexts. As an intelligent organism behaves, it can use the results of that behavior to learn and guide future actions. Yet, the world may be noisy, chaotic, and, in ways, constantly changing. In order for an organism to use the results of past experiences to inform behaviors in a new context, it must in some way generalize similar past experiences to determine how those past experiences should inform behaviors in that new context. That is, it must have a way of considering the ways in which an experience is similar to a context the organism had previously experienced. What an organism considers similar will have a non-trivial impact on how that organism behaves. In Shepard's "Toward a Universal Law of Generalization for Psychological Science," he describes an example of a circumstance where a bird may benefit from generalization:

> a bird that ingested a caterpillar bearing particular coloration and markings and found it delectable or sickening, must decide whether another object of more or less similar visual appearance is of the same natural kind and should therefore be seized or avoided, respectively (Shepard, 1987).

This quote describes how the way that the bird *generalizes* is critical to predicting its *behavior*– future behaviors are not just determined by the birds ability to discriminate the coloration and markings of the caterpillar. That organism's generalization may be described in terms of how it characterizes *similarity* between stimuli or concepts. Pavlov showed that dogs would salivate not only in response to the bell that preceded food, but also bells of similar pitch, salivating more as the bell's pitch approached that of the conditioned bell (Pavlov, 1927). This finding motivated many researchers to

develop methods for describing gradients of generalization from empirical data using both animals through operant conditioning (Guttman & Kalish, 1956) and humans through identification learning (Shepard, 1957, 1958, 1980). These characterizations quantified the relationship between independent variables, like sound frequency and light wavelength, and dependent variables, like same-different judgments or tapping frequency, providing hints at how organisms generalize stimuli.

Yet, our ability to quantify these relationships does not imply that they need be universal, nor invariant across stimuli, sense, nor species. Similarity between tones of different octaves (Shepard, 1982; Blackwell & Schlosberg, 1943), between hues at either side of the visible light spectrum (Shepard, 1965), and between shapes with different symmetries (Shepard & Farrell, 1985) all exhibit non-monotonicity with respect to independent variables.

Shepard explored ways to model and formalize the generalization data in terms of distance metrics between points in an abstract, psychological space (Shepard, 1987). The method he proposed assumed the existence of an n-dimensional psychological metric space for which there may exist an invariant monotonic function that uniquely maps distances between points in that space onto similarity data. Given a specified space, one can attempt to find this function numerically by iteratively shifting elements in a projecting matrix, in the direction of steepest descent in error between distances in the projected space and measures of dissimilarity, until the distance metric constraints are met to a tolerance (Kruskal, 1964; Shepard, 1962). The resulting matrix describes a function which, once plotted with similarity data points against distances between points, describes a gradient of similarity. This method provided a way to approximate similarity of stimuli in terms of distance metrics in an abstract psychological space.

The methods of Shepard and colleagues construct high dimensional vector spaces, within which points signify the experience of particular stimuli. Yet, these methods imply no claims concerning how such a space may be biologically implemented, or to what extent the characterization of a few gradients of similarity may describe the generalization of the organism ecologically. A very productive approach to conceptualizing how a psychological space may be implemented has been motivated by the application of continuous dynamical systems (Spivey, 2008). With this approach, the experiences of a cognitive organism are jointly encoded as points in a high dimensional vector space (Churchland, 1989). Much like in the Shepard case, points in this space are considered to be more or less similar by a distance metric, creating overlapping regions of conceptual similarity at many scales (Gärdenfors,

2000). Over time, mechanistic cognitive processes (Churchland, 1996) cause sequences of mental states to form trajectories in vector space (Yoshimi, 2012a).

While there are many ways to quantify similarity of neural dynamics, it is difficult to use those quantifications to explain the behavioral data of category formation, particularly in complex tasks incorporating uncertainty. One such complex task is the sexing of day old chicks at a farm for commercial egg production—a task studied in detail by Biederman and Shiffrar (Biederman & Shiffrar, 1987). For these farms, it is essential to separate male chicks from female chicks as quickly as possible, but the appearance of the genitals on the chicks are highly variable, with many possible arrangements, making this task very difficult to perform accurately. Maximum accuracy of 99.4% is achieved by experts after an estimated 2 to 6 years of sexing experience at an average rate of 960 birds per hour (although accuracy levels over 90% are achieved after a few months). Biederman and Shiffrar ran behavioral experiments, comparing the photo-based chicken sexing performance of naive subjects before and after instruction. Naive subjects averaged 60.5% correct before instruction, and 84% after instruction. If they were tested again without instruction, the accuracy actually decreased by 5% to 54.1%, implying that their performance increase was a result of the instruction. The instruction described various features that are useful for identifying the sex of the chick, all of which are useful in different scenarios, yet ultimately ambiguous, like"usually, but not always, male genitalia are larger" or "[male chicken genitals] tend to look round and fullish like a bell or watermelon" or "[female chicken genitals] can look pointed, like an upside-down pine tree, or flatish" or that "sometimes either sex will appear to have double genitalia." The research of Biederman and Shiffrar shows how people are able to succeed at complex tasks even without knowledge of exact category boundaries, and that nuanced knowledge of categories in these complex tasks drives performance better than simpler, clearer category boundaries. This research also exhibits the graded structure of category membership, which is sensitive to details that may not be explicitly identifiable.

The relationship between neural dynamics and conceptual representation significantly constrain the foundations of cognitive science and phenomenology. In this chapter, I will argue for some philosophical implications of considering PNGs as carrying conceptual information. A characterization of the significance of neural dynamics, described in terms of a noisy regular spiking context, will be used to understand bursting dynamics— particularly in neocortex.

4.1 From Conceptual Vector Spaces to Conceptual PNGs

One approach to mental representation has involved the application of a continuous dynamical systems perspective (Spivey, 2008). From this perspective, currently active concepts in a cognitive system (or subsystem) are jointly encoded as a point in a high dimensional *conceptual vector space (CVS)* (Churchland, 1989). Nearby points in this space, according to some distance metric, are seen as representing similar conceptual states, allowing regions and manifolds within this space to capture more general concepts (Shepard, 1987). The evolution of mental states over time becomes a trajectory in this vector space (Yoshimi, 2012a), driven by mechanistic cognitive processes (Churchland, 1989). In general, the CVS approach has been very productive.

Past challenges to the CVS approach have come from above: from more abstract and symbolic characterizations of cognitive processing (Fodor & Pylyshyn, 1988). More recently, a challenge has arisen from below: from insights into the neural coding of information. There is increasing empirical evidence that, in at least some neural systems, relevant information is encoded in the spatiotemporal pattern of spikes produced by neurons in a given nucleus (Rolston et al., 2007; Madhavan, Chao, & Potter, 2007; Pasquale et al., 2008). While information may be carried by synchronous or coherent firing of neurons, as in *synfire chains* (Bienenstock, 1995), computational considerations have suggested that content may frequently be encoded in complex asynchronous patterns of spikes (E. Izhikevich, 2006). As previously discussed, these complex spike patterns have been called *polychronous neuronal groups* (PNGs).

The PNG approach to representation differs substantially from the CVS approach (W. St. Clair & Noelle, 2013). A PNG is a temporally extended pattern of discrete spiking events over a collection of neurons. It is not clear how such a pattern could be mapped to a point in a continuous vector space so as to preserve relevant aspects of similarity between representations. A PNG need not be oscillatory, so it does not make sense to extract features like frequency or phase to map a PNG into a continuous vector space. In Section 4.2, I will show how the PNG approach differs from the CVS approach with respect to conceptual transitions, similarity, and other issues of representation.

4.2 PNG Triggering and Conceptual Representation

Past methods have characterized PNGs in terms of the precise stimuli that evoked them. The conditions that activate a particular PNG can be formally described.

For example, in the previously introduced definition of a σ-triggered polychronous group, a PNG is referenced by the precise spike-time pattern that caused it to occur, comprised of precisely σ number of spikes. However, there may be many possible spike-time patterns that produce that same PNG. These different PNG evoking patterns can be anywhere between small perturbations of the original considered σ-trigger neurons and timings, to completely different patterns that happen to trigger the same PNG. For example, on reoccurrence of a trigger, STDP strengthens the synapses that participate in the evoked PNG. As a result of this strengthening, the number of action potentials required to cause the necessary responses in downstream cells can be fewer, resulting in the same stereotyped PNG dynamic from fewer stimulating neurons.

Additionally, multiple σ-triggered polychronous groups can refer to what is essentially the same PNG; a specific σ-trigger is sufficient, but may not be necessary, to produce a PNG. The conditions under which a given PNG appears may be characterized by its *trigger set*, which I formalize as follows:

Definition A *PNG*, π, is a spatiotemporal spike pattern in a neuronal network captured by a set of neuron-time pairs. The *trigger set* of π, τ_π, is the set of spike-time patterns that trigger the existence of π. Each spike-time pattern in the trigger set of π will give rise to π when presented in the absence of interfering spikes.

The definition of a PNG's trigger set highlights the fact that the activation of a PNG can be caused in multiple ways. Note that each element of a trigger set is, itself, a PNG, since it is defined simply in terms of the neuron-time pairs that trigger it. This fact allows us to formally characterize the potential causal relationships between PNGs.

Definition Given a PNG, π_0, and that $\mathbb{P}$ is the set of all possible PNGs in the neuronal network, then the *meaning* of π_0, μ_{π_0}, is a pair of sets: its *existential component*, $\mu^E_{\pi_0}$, defined as the set of PNGs that contain a non-empty intersection with an element of the trigger set of π_0, τ_{π_0}, or

$$\mu^E_{\pi_0} = \{\pi \mid \pi \in \mathbb{P},\ \exists\ \pi_t \in \tau_{\pi_0} : \pi_t \cap \pi \neq \emptyset\};$$

and its *causal component*, $\mu^C_{\pi_0}$, defined as the set of PNGs whose trigger set, τ_π, contains an element with a nonempty intersection with π_0, or

$$\mu^C_{\pi_0} = \{\pi \mid \pi \in \mathbb{P},\ \exists\ \pi_t \in \tau_\pi : \pi_t \cap \pi_0 \neq \emptyset\}.$$

By this definition, the forward-looking causal component of μ_{π_0}, $\mu^C_{\pi_0}$, includes any PNG for which π_0 contains some spikes that may contribute to the triggering of that PNG. This means that $\mu^C_{\pi_0}$ includes PNGs that may only be triggered by π_0 in the context of other spike-time patterns occuring in the network. The backward-looking existential component of $\mu^E_{\pi_0}$ includes any PNG that contains spikes that may contribute to the triggering of π_0. This means that $\mu^E_{\pi_0}$ includes PNGs that may only trigger π_0 in the context of other spike-time patterns occurring in the network. Thus, *context sensitivity is intrinsic to the meaning of a PNG.*

With this description of PNGs, we have a way of characterizing how two PNGs can have similar downstream effects. Two similar PNGs, π_0 and π_1, will have similar overlap in terms of what PNGs it may help trigger. If we can quantify the cardinality of meaning sets μ_{π_0} and μ_{π_1}, denoted $|\mu_{\pi_0}|$ and $|\mu_{\pi_1}|$, then we can quantify the amount of this overlap. Specifically, the similarity between two PNGs may be described:

Definition Given the existence of two PNGs $\pi_0, \pi_1 \in \mathbb{P}$, and their meaning sets are μ_{π_0} and μ_{π_1}, then their real valued *meaning set overlap* is the cardinality of the intersection of their meaning sets, divided by the cardinality of the union of their meaning sets, defined both existentially and causally, or:

$$\left\|\mu^E_{\pi_0} \cdot \mu^E_{\pi_1}\right\| = \frac{\left|\mu^E_{\pi_0} \cap \mu^E_{\pi_1}\right|}{\left|\mu^E_{\pi_0} \cup \mu^E_{\pi_1}\right|}; \quad \left\|\mu^C_{\pi_0} \cdot \mu^C_{\pi_1}\right\| = \frac{\left|\mu^C_{\pi_0} \cap \mu^C_{\pi_1}\right|}{\left|\mu^C_{\pi_0} \cup \mu^C_{\pi_1}\right|}$$

These definitions form the foundation of a new neural account of conceptual representation. One leading philosophical account of the grounding of conceptual meaning is *Conceptual Role Semantics* (Block, 1997). In brief, this theory characterizes the meaning of a representation in terms of the causal relationships surrounding it. Thus, the meaning of a representation depends on the representations that can cause it, as well as the representations that it can cause, eventually making contact with sensory-motor processes interacting with the world. The formal definition of the meaning of a PNG, μ_π, makes explicit and precise the causal relationships between PNGs, allowing them to act as grounded representations according to Conceptual Role Semantics (W. St. Clair & Noelle, 2013).

There are additional features of PNGs that make them attractive for theories of neural representation. For example, a special case of PNGs, called synfire chains, have been studied extensively as a unit of representation. Generalizing from this research, spatiotemporal spiking patterns have been argued to have the essential properties of conceptual representation, including stability, reproducibil-

ity, learnability, storage capacity, and even compositionality (Bienenstock, 1995; Abeles, Hayon, & Lehmann, 2004; Hayon et al., 2005). Indeed, there are natural mechanisms for composing PNGs, potentially producing complex compositional representations, simply by superimposing them, with multiple PNGs being simultaneously active. Using my definition of PNG meaning, the meaning of the union of two simultaneous PNGs will be systematically related to the meanings of the two individual PNGs, providing an approach to compositional semantics (W. St. Clair & Noelle, 2013).

My definition of PNG meaning is *not* intended to suggest that every PNG in a neuronal network corresponds to a psychological concept. Indeed, since any reliably produced set of spikes fits the simple definition of a PNG, even very short spike sequences are formally PNGs. The definition of PNG meaning, however, does provide a perspective on which specific PNGs might correspond to useful concepts. Specifically, a PNG should be considered as a useful unit of representation to the degree that it arises in many contexts (i.e., the cardinality of its μ_π^E is large) and it participates in the triggering of a wide variety of other PNGs (i.e., the cardinality of its μ_π^C is large). From this perspective, useful conceptual units are ones that appear relatively independently. Thus, every concept is captured by a PNG, but not every PNG encodes a concept.

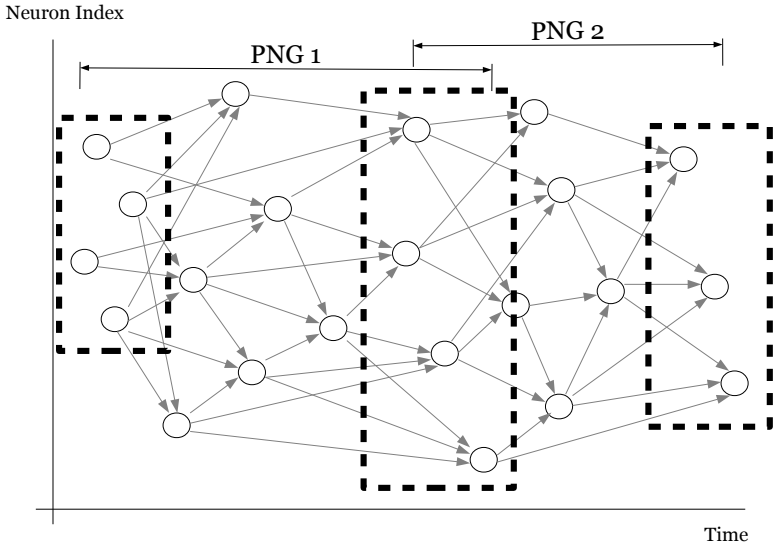

Figure 13: The vertical dimension specifies individual neurons in a group, and the horizontal dimension is time. Circles correspond to spikes generated by a particular neuron at a specific time. The arrows between spikes display causal effects between events. The dashed outlines contain groups of spikes whose existence is sufficient to trigger the following PNG.

To visualize the dynamics of PNGs, consider Figure 13. On the left, 4 neurons fire in a precise time pattern. This pattern is sufficient to trigger PNG 1. As PNG 1 unfolds, it includes a 4 spike sequence that is sufficient to trigger PNG 2. This second PNG can, itself, trigger further PNGs. If we now consider PNGs as encoding conceptual information, then PNG triggering dynamics can be seen as underlying conceptual transitions. Note that the triggers for PNG 2 do not need to be fully contained within PNG 1. We can imagine a different PNG 2 that is triggered with the aid of co-occurring contextual spikes in addition to those present in PNG 1. While this case is not depicted in the cartoon of Figure 13, in any complex neuronal network, the triggering of a PNG by multiple preceding PNGs will be the most common situation. This means that the activation of a PNG can be highly context sensitive. Also, recall that many PNGs can be simultaneously occurring without significant interaction, implying that PNGs can robustly coexist without undermining their heteroclinic stability (E. Izhikevich, 2006). This point about representational capacity is also supported by work on synfire chains (Trengove et al., 2013; Schrader, Diesmann, & Morrison, 2010).

4.3 Blended Transitions

One important aspect of a conceptual representation scheme involves the transitions from one conceptual state to another. Many researchers have suggested that concepts are best seen as regions in a CVS. Similarity is captured by distance metrics in this space. The space is also seen as a state space, through which dynamic trajectories unfold during cognitive processing (Spivey, 2008). The dimensions of the CVS are sometimes related to the instantaneous firing rates of groups of neurons.

In the CVS approach, a conceptual transition necessarily requires motion through an intermediate region of conceptual space, as portrayed on the left side of Figure 14. Along this dynamic trajectory, the system temporarily represents concepts that bear a steadily decreasing degree of similarity to the first concept, and a steadily increasing degree of similarity to the next concept. In this way, conceptual transitions inherently involve briefly activating intermediate and blended concepts. Extensive evidence has been gathered supporting CVS approaches (Shepard, 1987; Rolls & Tovee, 1995; Spivey, 2008), and a variety of psychological phenomena, including priming effects, have been addressed (Mirman & Magnuson, 2009; Cree, McRae, & McNorgan, 1999).

No intermediate representations need arise in the PNG approach. When concepts are taken to be encoded as individual PNGs, such conceptual transitions arise when the existence of one PNG

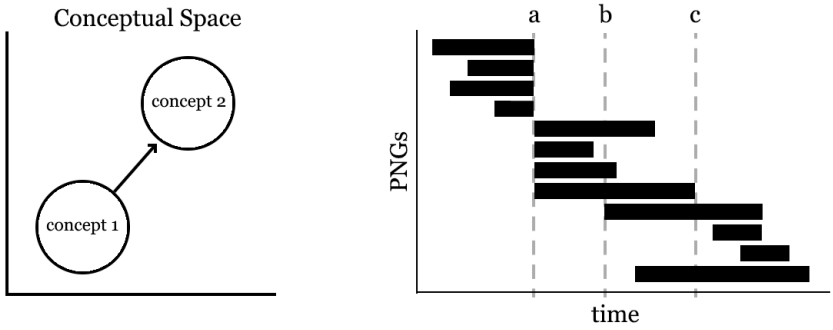

Figure 14: Left: A CVS, with concepts encoded as regions, with dimensions either being arbitrary or related to the firing rates of neurons. Conceptual transitions require motion through intermediate regions of conceptual space, which necessarily correspond to intermediate and blended concepts. Right: Both discontinuous and blended transitions between complex concepts, encoded as collections of PNGs, are shown. At time **a**, there is a discrete transition from an initial concept to a second one. Between times **b** and **c**, some PNGs from one concept are active alongside PNGs from another concept, capturing a blended conceptual transition during this period.

(often in the context of other, simultaneously active, PNGs) *triggers* the existence of another PNG, as caricatured in Figure 13. The unfolding of the initial PNG triggers the subsequent one, but at no point during this transition is there a spike-time pattern that is some sort of "blend" between the two spatiotemporal spike-time patterns that make up any of these PNGs. The only overlap between the PNGs is the sequence of spikes that trigger the second PNG, and that overlap can be vanishingly small compared to the extent of the PNGs. In this way, PNG representations allow for discrete conceptual transitions without the activation of blended concepts. This is possible because, with PNGs, the mechanisms driving conceptual transitions are decoupled from measures of conceptual similarity.

While spatiotemporal codes allow completely discrete transitions, they do not rule out transitions that activate intermediate or blended representations. Neuronal networks in the PNG framework may very well produce such blended states, particularly when the full conceptual state of the system is captured by multiple simultaneously active PNGs. As previously noted, since PNGs can be superimposed without much interference, the simultaneous unfolding of multiple PNGs would be a natural code for more complex compositional concepts. A conceptual transition from one complex concept to another might involve the deactivation of PNGs that make up the initial concept during the activation of PNGs that make up the following concept. If the initial concept PNGs all ended just as the following concept PNGs started, the transition would be discontinuous. If, however, the initial concept PNGs dropped out incrementally over time and the following concept PNGs began at a staggered schedule,

there would be an intermediate time at which some of the PNGs for the initial concept would be active alongside the PNGs for the following concept. This is a natural characterization of an intermediate or blended complex concept. An illustration of both discontinuous and blended transitions of these kinds appears in the right panel of Figure 14. In this way, the PNG framework allows for blended conceptual transitions, but, unlike the continuous vector space approach, it does not require them.

Finally, it is important to note that discontinuous jumps in conceptual space are possible in other spiking neuronal network frameworks, such as synfire chain representations (Bienenstock, 1995; Hayon et al., 2005). While I have used PNGs to highlight this difference between certain spike codes and CVS approaches, other frameworks could depart from continuous firing-rate models in similar ways.

4.4 Movement through an Emergent PNG Existence Space

The previously discussed $\phi\rho\psi$ formalization of emergence can be used to characterize conceptual transitions as emerging from neural dynamics. PNGs emerge from an underlying space with neural dynamics ϕ. ρ_π identifies the existence of sufficient spikes that form an element of the trigger set of π, where $\pi \in \mathbb{P}$ is emergent from the neural dynamics, ϕ. Each corresponding PNG space, on which ψ_π operates, propagates in a way similar to a row on the right in Figure 14. If we take ρ_π to be $\left\|\mu_\pi^C \cdot \mu_{\pi_0}^C\right\|$ or $\left\|\mu_\pi^E \cdot \mu_{\pi_0}^E\right\|$, where π_0 is constructed from propagating spikes in the underlying spike-time space that intersect with π, then the black bars in Figure 14 may become real values that may correspond to the extent to which a particular PNG has propagated, or its subcomponents are present in the network dynamics. Dynamics through such PNG spaces may be considered as a causally descriptive abstraction of the underlying neural space ϕ, with dynamics indicative of the functional properties of the systems they intersect.

A larger, unified set of PNG spaces may have dynamics that provide a view of cognitive neuroscience that uses representations that are decomposible into individual spikes, but are propagated in terms how they trigger other PNGs, providing an alternative to visualizing trajectories in a conceptual vector space.

4.5 Conceptual Representations and Dorsolateral Prefrontal Cortex

Neural firing patterns in the dorsolateral prefrontal cortex (DLPFC), thought to encode working memory contents, provide an interesting example of the variety of ways in which the PNG approach to

conceptual representation can inform our understanding of neural information processing.

4.5.1 Time Structure of Actively Maintened Representations in DLPFC

Mammalian brains represent a temporally rich and spatially complex environment, but are also simultaneously capable of representing abstract rules utilizing temporally stable representations. These representations are selectively preserved in working memory (Baddeley & Hitch, 1974), which neuroscientific evidence correlates with the activity of cells in the DLPFC (P. Goldman-Rakic, 1995). Representations are described as being *actively maintained* while in use, capable of preserving information for time scales much longer than dynamics occurring elsewhere in the brain. While the precise mechanisms of their active maintenence are not exhaustively known, they are thought to include topographically reciprocal synaptic projections between DLPFC, basal ganglia, and thalamus (G. E. Alexander, DeLong, & Strick, 1986), along with dopaminergically induced hyper-sensitivity (Durstewitz, Seamans, & Sejnowski, 2000) leading to the formation of attractors within interdigitated stripes of DLPFC (R. O'Reilly & Frank, 2006; Rougier, Noelle, Braver, Cohen, & O'Reilly, 2005).

Past approaches to modeling DLPFC representations have done so with the assumption that representational content is encoded by a distributed population of active neurons, where different amounts of activity over those neurons correspond to different possible contents of a particular kind. Here, an alternative encoding scheme is examined, polychronous neuronal groups, where the precise timing of neuronal action potentials can provide additional detail to representational contents, triggering potentially distinct responses throughout more posterior regions of cortex (E. Izhikevich, 2006). With this representational lens, implications are considered regarding essential prefrontal cortex functions, including the ability to update stored information, robustly maintain information, and drive other processes in the brain (R. C. O'Reilly, Noelle, Braver, & Cohen, 2002), though without excising potentially essential temporal information. I suggest that polychronous information may underlie processes that are otherwise driven by mean activity dynamics, bearing information accessible upon a representations propagation to more temporally sensitive cortical subregions.

4.5.2 Neural Correlates of Working Memory

The study of cognition has long made the distinction between controlled, effortful actions, and automatic, effortless actions (James, 1890). Classically, the Stroop task well characterizes this distinction

(Stroop, 1935). In the Stroop task, subjects are presented with color word stimuli with a colorized typeface, and are asked to identify aloud either the word or the color of the ink. In congruent tasks, the ink and the color word match, resulting in a faster time to name the correct response, while for incongruent trials the ink and the color word differ, resulting in a much slower time to name the correct response when asked to name the color of the ink, as reading the word evokes a stronger response than identifying the color. Congruent trials are described as being automatic and effortless, while incongruent trials require great control, requiring the effective inhibition of the otherwise more dominant word reading response. Critically, those with damage to their DLPFC are much slower still to respond on incongruent trials, and produce more errors (Stuss, Floden, Alexander, Levine, & Katz, 2001). Baddeley and Hitch distinguished the need for a working memory system with properties distinct from the strictly described short-term and long-term memory mechanisms, with contents that could be updated according to task demands (Baddeley & Hitch, 1974). While there is consensus on the need for something of this vague cognitive scope, its neuronal implementation remains unresolved.

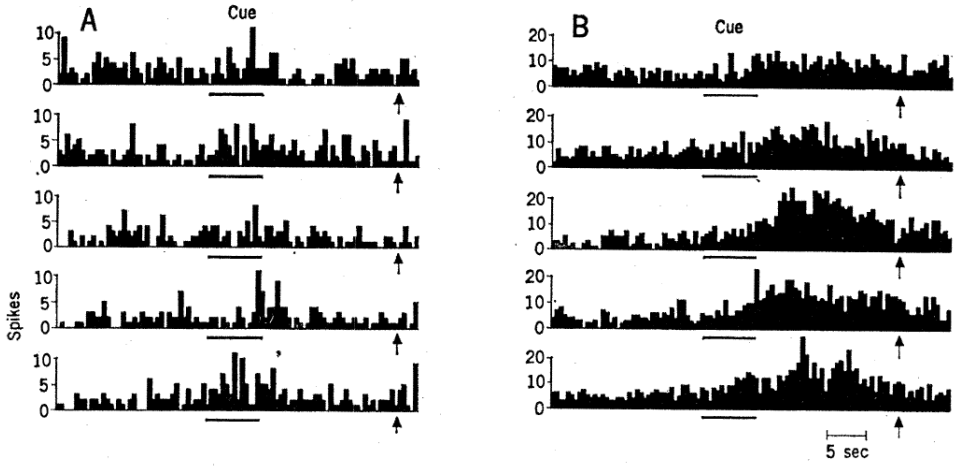

Figure 15: In this figure, spike histograms from recordings of DLPFC in rhesus macaques during the delayed response task are plotted. On the left, a cell is plotted that showed no correlation with the task relevant maintenance, as compared to the cell on the right, which shows maintained activity during the delay period across trials (Fuster et al., 1971).

Prefrontal cortex is suspected to play an important role in the ability to organize thoughts and actions relevant for internal goals. Indeed, one of our first neural correlates of working memory is from DLPFC (Fuster et al., 1971). Electrodes inserted into the DLPFC of rhesus macaques to study a delayed response task showed activity that correlated with the presence of the obscured cue during

the delay period. In the delayed response task, monkeys observed an apple being placed within one of two boxes, and the box was covered with a lid. Then, after a delay, the monkeys had to select the box that contained the apple to receive the reward. During this delay period, the mean activity of recorded neurons was maintained above baseline (Figure 15). This finding has been generalized to show that the activity of cell groups in DLPFC during the delayed response task can be specific to different regions in the monkey's visual space (P. S. Goldman-Rakic, 1987). Similar effects can be shown for the encoding of abstract rules and the existence of particular objects (Wallis, Anderson, & Miller, 2001). Indeed, since failure to maintain this activity is associated with behavioral errors (Funahashi, Bruce, & Goldman-Rakic, 1989), evidence suggests that activity in DLPFC may constitute the cellular basis of working memory (P. Goldman-Rakic, 1995).

4.5.3 Physiology of a Working Dorsolateral Prefrontal Cortex

While there is abundant evidence that activity similar to what is expected from a working memory system is occurring in DLPFC, it remains to be shown precisely how these representations are updated and maintained. Before we characterize DLPFC's embedded dynamics, however, we should consider its internal context.

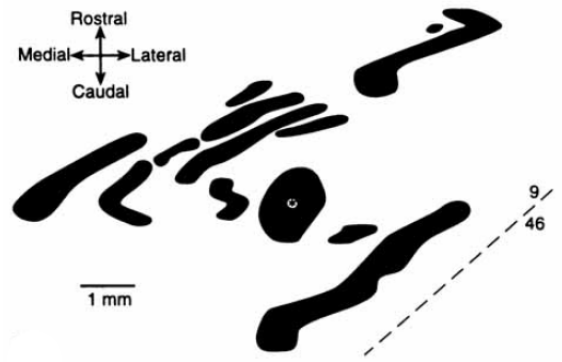

Figure 16: In this figure, anterograde and retrograde tracers were injected at the point of the asterisk. The black shapes correspond to regions within which activity is mutually excitatory. These shapes are spatially interdigitated with similar patterns of shapes, whose activation has an inhibitory effect. In this way, each interwoven stripe in DLPFC excites or inhibits neighboring stripes (Pucak et al., 1996).

DLPFC has a distinct pattern of internal excitatory projections, seen in Figure 16, forming stripes within a layer when labeled with retrograde and anterograde tracers (Pucak et al., 1996; Levitt, Lewis, Yoshioka, & Lund, 1993). There are many groups of interdigitated stripes, each with coherent columnar

structure across layers. Approximately three quarters of these long range intralaminar connections are monosynaptic (Henze, González-Burgos, Urban, Lewis, & Barrionuevo, 2000). Because of their proximity, depolarization induced GABAergic responses resultant from activity within one stripe can effect nearby stripes, forming a grid of mutually inhibitory patches. This unique connectivity pattern has been used to motivate models of DLPFC function where each stripe is seen as capable of storing distinct representations (R. O'Reilly & Frank, 2006). Activity in superficial PFC is relatively sparse compared to the rest of cortex, though activity in supragranular layers 2/3, which receive projections from thalamus, is non-sparse (Barth & Poulet, 2012). This carries the implication that the sparsity of superficial DLPFC is the result of the cellular and physiological mechanisms distinct to the region.

Yet this description is not sufficient to explain how novel items can be stored in these stripes. One significant component is the existence of NMDA receptors in the synapse (Lisman, Fellous, & Wang, 1998). The impact of NMDA receptors is dependent on the membrane potential of the postsynaptic neuron. If the postsynaptic neuron has recently achieved an action potential, and its membrane potential is high, then NMDA receptors proportionally become unblocked. When NMDA receptors are unblocked, they serve to reduce the rate of repolarization, allowing the cell to continue firing when it would otherwise require significant input to maintain that level of activity.

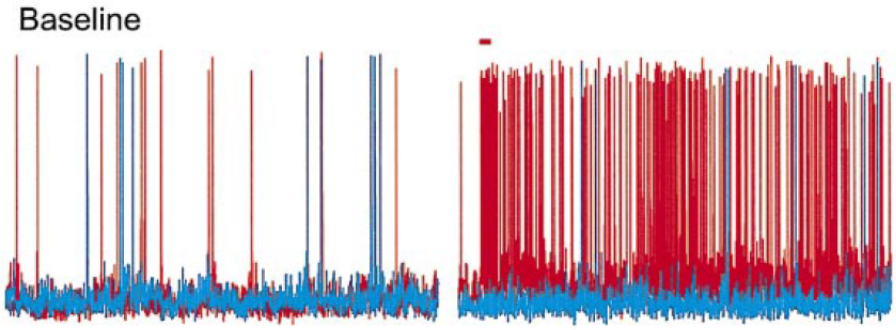

Figure 17: This figure shows the difference between high and low dopamine levels on target neurons (red) as compared to non-target neurons (blue) in a cortical model. The small horizontal red bar indicates the time and duration of the initial stimulus. On the left, we see the low dopamine case. On the right, we see the high dopamine case. In the presence of more dopamine, target cells maintain high firing rates, while non-target cells remain at baseline (Durstewitz et al., 2000).

Another critical component to the ability to update and maintain representations in DLPFC is the presence of dopaminergic signals from basal ganglia (Williams & Goldman-Rakic, 1995). Peaks in dopamine serve to amplify NMDA currents, and shift sodium and potassium currents toward hy-

perpolarization, while also increasing the speed and response of GABA activity (Durstewitz et al., 2000). The implication is that bursts of dopamine to PFC can serve to strengthen existing dynamics, making the maintenance more resilient, while also increasing inhibition, which diminishes competing representations (Figure 17). A reduction of baseline dopamine levels, like when outcomes do not meet expectations (Schultz, Dayan, & Montague, 1997), would cause a relative destablization of maintained representations through the reduction of NMDA currents and the quickening of repolarization. Dopamine also has a synapse specific impact on DLPFC neurons which reduces the strength of incoming signals without reducing excitation between stripes, helping prevent interference with what is being actively maintained (Gonzalez-Burgos et al., 2002). It also increases the excitability of many kinds of inhibitory interneurons, which can also help to reduce interference.

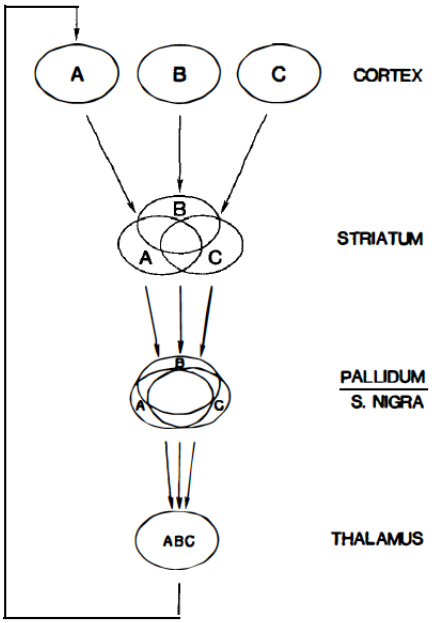

Figure 18: This figure portrays the convergent projective overlap of several functionally related cortical regions within thalamocortical loops. While there are recurrent projections, mapping back onto themselves, a single cortical region also facilitates activity in similar functional cortices (G. E. Alexander et al., 1986).

Recurrent connections can also serve to help maintain active representations (Zipser, Kehoe, Littlewort, & Fuster, 1993). Beyond the recurrent connectivity within stripes, DLPFC also has reciprocal topographic projections with striatum and thalamus, where the returning projections significantly overlap with the neurons that initially projected, portrayed in Figure 18 (G. E. Alexander et al., 1986). It is important to recognize that activity in thalamocortical loops need not project perfectly

back onto the cells that were initially active. Nearby projections from PFC to striatum have 40% - 80% projective overlap with cells as much as 5 millimeters away, approaching an average of 20% projective overlap 10-50 millimeters away (Averbeck, Lehman, Jacobson, & Haber, 2014). This particular aspect of thalamocortical loops has been argued not only to facilitate active maintenance (R. O'Reilly & Frank, 2006), but has also to enable representations to become associated with new information indirectly, which has been shown to allow networks to successfully generalize knowledge to novel stimuli (Kriete, Noelle, Cohen, & OReilly, 2013).

4.5.4 Temporality of DLPFC

To what extent is precise spike-timing information relevant for DLPFC? There is evidence for spatiotemporal representations throughout cortex: in auditory cortex with millisecond precision (Lu, Liang, & Wang, 2001; Kayser, Logothetis, & Panzeri, 2010); in visual cortex (Engel, König, Kreiter, Schillen, & Singer, 1992) and with visual information in thalamus (Reinagel & Reid, 2000); in motor output and premotor neurons (Fu, Flament, Coltz, & Ebner, 1995); and even in hippocampus (Huxter, Burgess, & O'Keefe, 2003). Whether or not activity in DLPFC *itself* contains temporal information, the systems that it is interacting with do appear to contain temporal information.

Mechanisms of active maintenance of representations in DLPFC appear to be preserving of activity, but it is unclear how information maps to and from these neurons. While recordings from delayed response tasks clearly exhibit sustained activity with respect to some feature dimension (as in Figure 15), the firing rates of these neurons vary throughout their maintenance. With such variety, it seems unlikely that content is significantly preserved through combinations of precise rates of firing. The sparseness of coding and sustained activity corresponding with particular features, however, seems to suggest a strong spatial significance to what is maintained. This notion is reinforced by the prevalence of monosynaptic projections between stripes. Interestingly, these lateral connections are unusually slow, sending potentials at 0.14 meters per second, around 400 times slower than average, slower than the slowest unmyelinated axons (Henze et al., 2000). If there were spatiotemporal structure in the stimulation of DLPFC, it is unclear if the lateral excitatory connections between stripes would preserve it beyond the temporal dynamics of their mean activity.

There can be temporal information in mean activity, however. Event relative time-stamp information emerges in DLPFC, akin to timing signals used in reinforcement learning explanations of dopamine

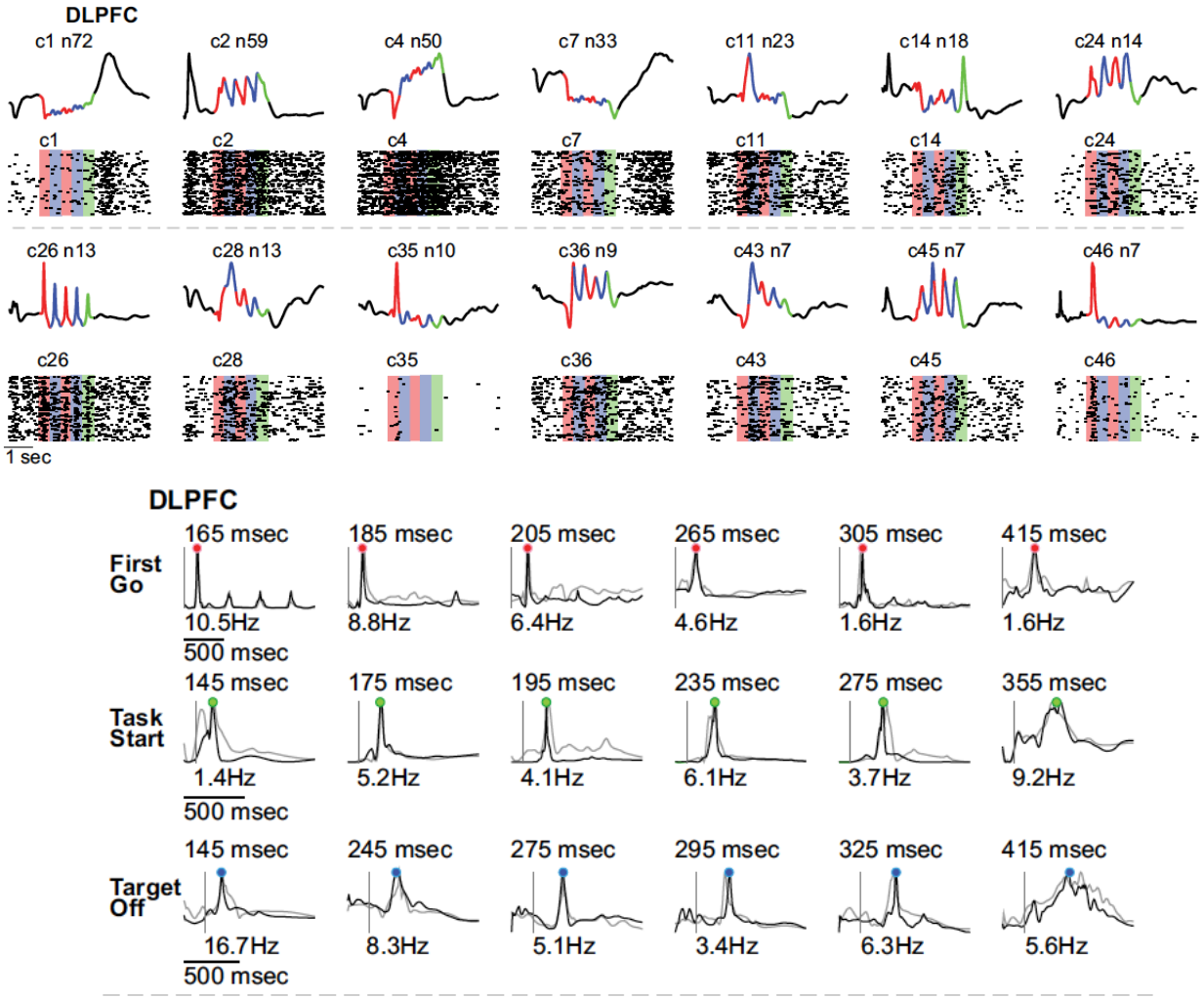

Figure 19: *Above:* Average response profiles of selected neural clusters in DLPFC during the sequential saccade task. Red bars indicate the presence of the go cue, blue bars indicate the stop cue. At the end of the trial monkeys still fixated vigilantly until they received their variably timed reward. *Below:* This figure shows evidence for time-stamp profiles for single DLPFC neurons. (Jin et al., 2009).

bursts (Schultz et al., 1997), which can be observed in monkeys that are trained on the sequential saccade task. In this task, monkeys are trained to saccade to four sequential cues presented at gaussian-distributed intervals, and are rewarded with juice if they successfully fixate (Jin et al., 2009). From multi-electrode arrays, neurons can be observed that have sharp temporal peaks at points distributed across times after the task begins, portrayed in Figure 19. Monkeys performed this task nearly automatically, having performed 800 trials a day for 3 years across studies. This provides insight into event-relative timing emerges, and how dynamically DLPFC can represent information. In (Jin et al., 2009), precisely timed saccades appear to be spatially coded events, with peaks at precise times.

This suggests that systems guiding DLPFC are sensitive to precise temporal information, however that information is implied to be spatially coded in a way akin to identifying the brief occurrence of an event.

4.5.5 Regular Significance of Bursting Dynamics

Izhikevich probed randomly connected networks to quantify the number of supported σ-triggered PNGs to show their ubiquitous emergence, where $\sigma = 3$—even in the presence of background noise (E. Izhikevich, 2006). The stability of the PNG, in this experiment, was characterized by its imperviousness to interruption, tested by providing a context of uniform noise. The significance of a particular stimulation in this paradigm is taken to be what emerges out of repeated stimulations an indication, each measured independently. However, if we were to change the stimulation pattern, different heteroclinic chains may emerge. The size of the neighborhood where that stimulation can be modified without changing the way the original PNG from propagates can be seen as indicative of the stability of that PNG as a heteroclinic channel.

The definition of a PNG from Section 4.2 relies on the notion of a spike's *interference*. The causal meaning, μ^C, of a particular spike, generated at a relative time, can be approximated in the $\sigma-triggered$ quantification case in terms of how adding that spike-time pair to the $\sigma-trigger$ modifies the original quantification. If including the spike-time pair modifies the number of 3-triggered PNGs—or the manner and extent of their propagation—then the that spike-time pair can to be said to *interfere* with the 3-triggered PNGs altered by inclusion. In this way, a spike's interference is inherently in terms of the network structure in which it is embedded. As a neuron has more and stronger connections, the interference of its spikes are likewise enhanced.

This test of interference assumes a context of regular spiking neurons. In this regular spiking context, a spike is only considered as occurring at a particular relative time. However, in the brain, regular spiking neurons, which tend to fire at no more than 50 Hz, are intermixed with intrinsically bursting and rhythmically bursting neurons, which can fire between 300 and 600 Hz (Steriade, 2004). How can we describe the significance of a particular neuron's bursting in context of a regular spiking environment? A particular neuron's firing represents a *constraint* on what can emerge from the neurons that it projects to. If that neuron projects to one that exhibits regular spiking, then the impact of its spiking is over after just a few milliseconds, with less and less residual membrane potential as you get

further away in time (Cessac et al., 2010). This implies that a single spike can only interfere or apply constraints within a short time of its receipt. A neuron firing again ostensibly renews that constraint. As a neuron increases its firing rate, its constraints are applied to more time windows with less time between those windows (leading to more interference). As the constraint time windows get closer together and overlap, so does the likelihood that it will interfere with co-occurring propagations. As a neuron increases its firing rate, it consistently reapplies the constraint implied by the neuron, creating a context where what emerges must be increasingly harmonious with its activity.

A bursting neuron, which can fire as fast as 900–1000 Hz in thalamus, applies its constraints in an overlapping manner, in such a way that the likelihood of its interference during the time of activity becomes much stronger. This near guarantee of interference can also be seen as a guarantee of facilitation, in that all PNGs which emerge in the context of that neuron's bursting should either intersect with that neuron or be in some way harmonious with it. When considering the significance of a population code, seen as a distribution of firing rates over a pool of neurons, that particular population code will constrain what emerges in its embedded network without fully prescribing what emerges. In this way, analogues to population codes, like blood oxygenation in fMRI or the electric field potentials measured by EEG and ECOGS, can be correlated with hypothesis of information flow—even in the absence of a more detailed theory of neural coding.

While the notion of interference is in terms of the obstruction of an otherwise capable PNG, the way that a spike changes a co-occurring propagation may also be considered positively—the significance of spike can be seen as an unrelated distractor to a propagation or a computationally significant contributor. In this way, the impact of a spike's interference can itself propagate, potentially allowing its residual impact to be observed at different scales. This could allow neurons with sparse firing rates to significantly contribute to the constraint of processes well between the time windows where the spike directly interferes.

The repeated activation of a single bursting neuron, projecting to regular spiking neurons, itself need not be sufficient to prescribe further propagation. However, that bursting neuron may be seen as providing a strong *constraint* on accompanied dynamics, leading to meaningfully related propagations.

To clarify the notion of constraint, consider this framing question: when a neuron fires repeatedly, as in a burst, how does it change its meaning set? Since a meaning set for a single spike contains all PNGs it intersects with, considering the meaning of that spike in isolation will yield the largest

meaning set possible for the neuron of interest, while the proportion of PNGs in that meaning set with sufficient triggers will be the smallest possible for that neuron.

When we include a second spike by that neuron, then the size of that new meaning set is less than or equal to the size of the meaning set of one of those spikes alone. This is because the meaning set for a single spike is at its maximum size, and taking the second spike as given excludes all PNGs from that original meaning set that it interferes with. The reduction of the size of the meaning set with spike inclusion would not be guaranteed if we considered two spikes from two different neurons, since in that case the meaning set could grow by the number of PNGs intersecting only with the new neuron, which may be more than the number of PNGs omitted by interference with the new spike (even while the number of sufficiently triggered PNGs in their meaning set may increase). If the inclusion of the second spike increases the number of sufficiently triggered PNGs in the meaning set, then the original meaning set is ostensibly trimmed by the interference of those spike-times, since a PNG is defined in terms of propagation in the absence of interference. Furthermore, the assumed silence in the time between spikes also excludes PNGs which depended on a spike occurring during that time.

To restate, as a neuron spikes a second time, the meaning set created by the original spike may be seen as being trimmed by the interference of the new spike, while the number of sufficiently triggered PNGs may increase. The increase in sufficient triggers further changes the considered meaning set by providing additional interference as the spikes in that chain successfully propagate. As more spikes are added by the same neuron, the meaning set constructed by the full spike sequence can only get smaller, while the number of sufficiently triggered PNGs within that meaning set may grow. If we now consider the trigger set overlap of the PNGs contained in the meaning set, where complete overlap implies a sufficient trigger, then the degree of overlap of meaning set PNGs may increase as more spikes are added by the same neuron. If we were to quantify the number of spike-times that would need to be included for a particular PNG to be sufficiently triggered, then that number may decrease as more spikes are added by the same neuron. This reduction in required spikes constitutes the facilitation of that PNG. The landscape of facilitation in the PNG space of the meaning set describes the way in which a neurons spike shifts the probability of particular outcomes. As a neuron repeatedly fires, the interaction between the facilitation of PNGs in its meaning set and the interference of the sufficiently triggered PNGs contained in that set, has the impact of amplifying the reduction in size of the meaning set while increasing the likelihood of particular outcomes within that now smaller meaning set. The

way in which this occurs constitutes a particular neurons constraint on the network.

4.5.6 Regular Significance of Superficial DLPFC

Let us now consider bursting neurons in superficial DLPFC. While spatiotemporal coding may not be propagated laterally, it is possible that which *particular* neurons become active are dependent on spatiotemporal interactions. Recall that while superficial DLPFC is considered sparse, its supragranular DLPFC input neurons are *not* considered sparse (Barth & Poulet, 2012). Since these neurons receive their input from thalamus and sensory cortex, it may be reasonable to assume they could exhibit spatiotemporal correlations. Furthermore, because of the partial projective overlap in striatum, the structure of recurrent patterns could spread out over DLPFC as feedback progresses in a sequentially meaningful way.

The sequential meaning of partially overlapping thalamocortical feedback loops has successfully been shown to aid in the generativity of mean activity representations in DLPFC (Kriete et al., 2013). The difference in this circumstance is that the neurons that are being sparsely activated are ones *harmonious with the spatiotemporal codes that intersect the participating thalamocortical loop*. Because of this, the active maintenance of sparsely coded neurons in DLPFC could effectively drive the re-activation of a PNG by sufficiently overlapping with that PNGs trigger set. This occurrence is made more likely in this circumstance because the neurons intersecting with the trigger sets of PNGs participating in the DLPFC thalamocortical loop are the neurons that were preserved through dopaminergically induced active maintenance. The high levels of activity of trigger neurons, coupled with the inhibition of competing triggers, acts as a figurative *fire hose* of activation sufficient to enable meaningfully related processes elsewhere in cortex. Continuing with this metaphor, the spatiotemporal representations of cortex are the resulting dynamics of bursting neurons selected to facilitate particular PNG families, while being partially agnostic to which particular PNG gets triggered and the timing of its initiation.

This approach could harness the compositional nature of PNG representations (W. B. St. Clair & Noelle, 2015). As the DLPFC coding includes more features, the family of PNGs that can be invoked gets more specific. As the coding includes less features, the PNGs become increasingly underdetermined. As a result, what PNGs get triggered become more contextually defined. Following this explanation, DLPFC coding enables selective outcomes, while allowing for more detailed nuances in

terms of the precise impact of what has been actively maintained.

Brains exist in a spatiotemporally complex environment, and that complexity carries over into neuronal representations. We are capable of selectively attending and preserving critical features of those representations in working memory. This working memory can actively maintain, update, and forget current representations, while also being capable of driving processing elsewhere in the brain. Neural correlates imply that activity of cells in dorsolateral prefrontal cortex correspond to the dynamics of working memory. I suggest that the contents of DLPFC representations and bursting neocortical cells may be developed in harmony with the spatiotemporal representations of a regular spiking context, and may be best thought of in terms of their intersection with the trigger sets of the PNGs that they may either facilitate or inhibit.

4.6 PNGs, Linguistics, and Iconicity

As a final example of how the PNG approach to conceptual representation can influence our understanding of cognitive processes, I offer a PNG description of the development of language perception. In linguistics, a phoneme is defined as "the smallest segmental unit of sound employed to form meaningful contrasts between utterances" (IPA, 1999). Over time, the cochlea is a constant source of spike signals. At a basic level, each inner hair cell corresponds to a physical location in the cochlea. Movement of the hair cell correlates with release of glutamate in its synapses, generating a membrane potential in receiving neurons. These neurons send signals through the cochlear nerve to the cochlear nuclear complex (CNC), where they are received in a tonotopically organized fashion. These hair cells produce action potentials at the peaks and troughs of their oscillation, exhibiting a transient firing rate correlating with sound frequency. This implies that spikes in the spike patterns that emerge in the CNC correlate with the frequencies of a fourier decomposition of the audio signal. Formants are often used to describe the peaks in this spectral space that are generated by the human voice (Fant, 1960). Each phoneme has a distinct formant pattern given a particular speaker. In Figure 20, we see two main formants characterizing the American English vowels i, u, and a. Formants indicate frequency patterns that are most distinct to those vowels.

An infant hearing a phoneme can be thought of as the triggering of a polychronous group by ϕ_{CNC}, which causes a chain reaction in the network. As an infant repeatedly hears a phoneme, the chain reaction it causes strengthens. Development of phoneme representations is language specific. An infant

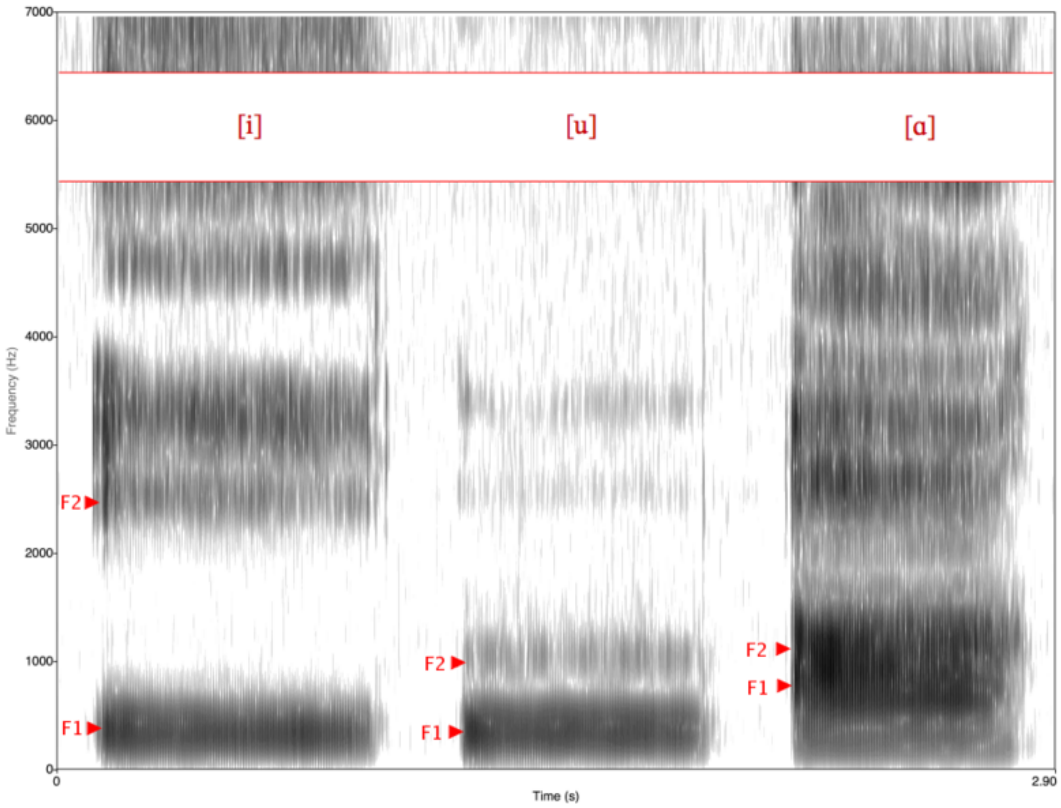

Figure 20: Spectrogram of the American English vowels [i, u, a] as pronounced by a native from Louisiana demonstrating the frequencies of the first and second formants.

that is repeatedly exposed to the Estonian language will develop increased phonemic awareness of the phonemes more common in Estonian–a different phonemic awareness than that of an infant exposed to Finnish (Cheour & et al, 1998). Over development, perception becomes increasingly categorical. By 11 months of age, an infant shows signs of an increase in native-language consonant perception, and a decline in foreign-language consonant perception (Kuhl, 2004). As a phoneme becomes learned, its PNGs become strengthened. However, once those neurons strengthen the connections relevant to its meaning, it weakens connections irrelevant to its meaning. Because of STDP, if a post-synaptic neuron fires just *before* its perception of a spike in a synapse, then it *weakens* the strength of that synapse. Inherently, signals that are repeated and have systematic correlative structures, like formants, will have beginnings of PNGs strengthened. Neurons that are participating in those PNGs will weaken their weights for frequency combinations that are not contained in those PNGs. As those phoneme PNGs become well learned, their meaning starts to become more significant. This is because the PNGs will become more easily triggered, and will be perceived in more circumstances, since the PNGs that have a trigger set overlapping with the triggered PNGs with have a part of their trigger set strengthened.

Figure 21: A See n' Say children's toy manufactured by Fisher-Price. The child spins the arrow, which points to an animal. Then, the child pulls the chord, which spins the arrow as a voice says the name of the animal, and imitates the sound it produces. The See n' Say provides an example correlated visual and auditory information which can be used to strengthen PNGs that otherwise have little else to correlate with.

Latent PNGs with trigger sets dependent on these strengthened components may now, themselves, be more easily triggered.

Since the mechanism that binds the formants and phonemes together is general, it can be used to learn multi-modal associations. As PNGs become stable, their meaning can combine with the meaning of other PNGs that are temporally co-occurring. In Figure 21, we see an example of a See n' Say, a children's toy that specializes in combining visual and auditory information. By seeing a cow, and hearing 'moo', a distinct pattern of auditory PNGs – those triggered by 'moo' – co-occur with a distinct pattern of visual PNGs–those triggered by the cartoon depiction of a cow. Given the small-world structure of the brain, these PNGs will develop to use as much information as possible, likely combining auditory and visual information. Over time, this multi-modal strengthening of PNGs could potentially yield a multi-modal PNG representation of a Cow, which contains many PNGs which may be triggered corresponding to a Cow's features–implying the formation of an internal representation of a Cow that is constructed from other PNGs, yet when triggered, has the *meaning associated with the experience of a Cow.*

In Section 4.2, I describe how PNGs have context dependent meaning. Consider a sequence of phoneme PNGs triggered in succession, $\pi_1 - \cdots - \pi_{10}$. Using the definition of meaning sets, the meaning of phoneme PNGs may be reconsidered as they become triggered, shaping the structure of the final PNG that is triggered by the full sequence.

A PNG sequence need not be triggered in full to contribute to a sufficient trigger. PNGs trigger as soon as their trigger set is satisfied. As we read the example "I am hungry and I am happy,"

we understand the meaning of "I am hungry" before we combine it with the meaning "I am happy." Indeed, spacing them into seperate successive sentences "I am hungry" ... "I am happy" does not provide the same context for meaning as "I am hungry and I am happy." By considering the properties of PNGs in a stable PNG existence space, we find what is fundamental in linguistics to be decomposible in cognitive neuroscience.

5 How Neural Firing Patterns Depend on Network Topology

The causal meaning of a PNG and the constraint of individual neurons, as described in the previous chapter, are defined in terms of the dynamics of the network of interest. How does a given network architecture impact the way that particular neurons constrain dynamics? Particularly, how do the neural firing patterns emerging from fully recurrent networks relate to changes in the pattern of network connectivity and conduction delay across and between clusters of neurons? I approached this question empirically through a large parametric study which simulated clustered network topologies of cortical excitatory neurons with inhibitory interneurons. Network architecture was algorithmically constructed from every possible combination of 6 independently varied parameters with 5 settings each to model a total of $5^6 = 15,625$ different network structures. These included 3 intercluster parameters and 3 intracluster parameters: minimum conduction delay, range of conduction delay, and connection density. Each network was stimulated for 1000 ms at various frequencies in a noiseless environment by an arbitrary rate code. The firing patterns evoked by each network are then described and analyzed to glean the distinct properties of each parametric variation.

From this massive grid search of the space of network structures, networks were found to be either: supercritically excited, with a vast array of explosive firing patterns (most common); subcritically quiescent, where no firing patterns were evoked despite tonic stimulation; or, metastably active, where tonic stimulation yielded likewise tonic firing within the network. Whether a particular network was found to be supercritical, subcritical, or metastable ultimately depended on its distinct combination of parameters.

5.1 Simulation Methodology

Each recurrent network consisted of 800 excitatory regular spiking neurons, modulated by 200 inhibitory fast spiking neurons, simulated using the Izhikevich simple neuron model (Equation 2.2) with

uniform parameters for each neuron type (E. Izhikevich, 2006). Excitatory neuron models used parmeters $a = 0.02$, $b = 0.2$, $c = -65$, and $d = 8$. Inhibitory neuron models used $a = 0.1$, $b = 0.2$, $c = -65$, and $d = 2$ (E. Izhikevich, 2003). These neurons were evenly divided into 5 *clusters*, each containing 200 excitatory neurons and 40 inhibitory neurons, emulating the 5:1 ratio of excitatory to inhibitory neurons observed in cortex. Each excitatory neuron projected to both excitatory and inhibitory neurons within its home cluster, but projected only to other excitatory neurons outside of its home cluster and never projected to inhibitory neurons outside of its home cluster. Each inhibitory interneuron neuron projected only to excitatory neurons within its home cluster, and never to neurons outside of its home cluster and never to other inhibitory neurons. Each spike through an excitatory synapse resulted in a $5mV$ change in the membrane potential of the receiving neuron. Each spike through an inhibitory synapse results in a $-4mV$ change in the membrane potential of the receiving neuron. All synapse strengths were fixed for the entirety of the experiment.

The number of connections that each neuron connected to within its home cluster was set by the parameter *Internal Connection Density*, while the number of connections that each excitatory neuron connected to outside its home cluster is set by the parameter *External Connection Density*. The minimum action potential propagation delay along axons which projected within the source neuron's home cluster was set by the parameter *Internal Delay Minimum*. The range of those delays, with uniform distribution starting from the Internal Delay Minimum, was set by the parameter *Internal Delay Range*. The minimum action potential propagation delay along axons which projected outside of the source neuron's home cluster were set by the parameter *External Delay Minimum*. The range of those delays were set by *External Delay Range*. Each neuron's non-repeating projection targets were determined uniformly at random.

The full 5 cluster network was stimulated by a 200 neuron input cluster which projected to the remainder of the network with a number of connections determined by the External Connection Density. This input cluster has no internal connections, serving as the only source of activity to the network.

For each condition to be tested, a new network was generated with the parameters of interest. This network was then stimulated according to a uniformly distributed rate code which scaled the expected number of spikes to be generated for each neuron's interspike intervals. These interspike intervals were sampled from a poisson distribution with λ mean and variance, where λ was set by the *Stimulation Frequency* parameter. Each network is stimulated in this way for 1000 milliseconds, with 1 millisecond

temporal resolution (computed in two steps of 0.5 milliseconds for numerical stability). This process was repeated for every unique combination of parameters in Table 1.

Parameter Conditions:	1	2	3	4	5
Internal Connection Density	5	8	11	14	17
Internal Delay Minimum	1				
Internal Delay Range	0	4	8	12	16
External Connection Density	5	8	11	14	17
External Delay Minimum	1	7	13	19	25
External Delay Range	0	4	8	12	16
Stimulation Frequency λ	5	15	25	35	45

Table 1: Parameter values explored in the analysis. Connection densities represent the number of projections per neuron. Delays are listed in milliseconds. Frequency of stimulation is in Hz, and corresponds to the maximum possible frequency within its rate coded stimulation.

An example of a network connectivity pattern used in the investigation is portrayed in Figure 22, where each point implies the existence of a projection from the horizontal to the vertical, and color implies that connection's action potential propagation delay (ranging from black to copper from 1ms–25ms. In this example, each neuron projects to 14 neurons within its cluster with a delay ranging from 1–9ms. These internal connections can be seen as rectangles along the diagonal, since the neurons are ordered with respect to their clusters. The last 40 neurons of each cluster were inhibitory, and hence only projected locally and never to themselves (made clear by the absence of connections in the top-right of each internal cluster). Furthermore, the inhibitory neurons were never distally projected to, shown by the vertical columns of white along inhibitory neuron indexes. Since no part of the network projected to the input driver, projections to neurons 0–240 are absent (shown by the white band on the lower portion of the graph). In this example, each neuron also projected to 8 neurons outside of its local cluster with a delay ranging from 7–25ms. This network had dense and fast local connectivity, but slower and more sparse intercluster connectivity. Compare this to a second example, shown in Figure 23, where clusters were less densely connected internally with longer delays, while being more densely connected externally with shorter delays. A priori, it was unclear how these different network

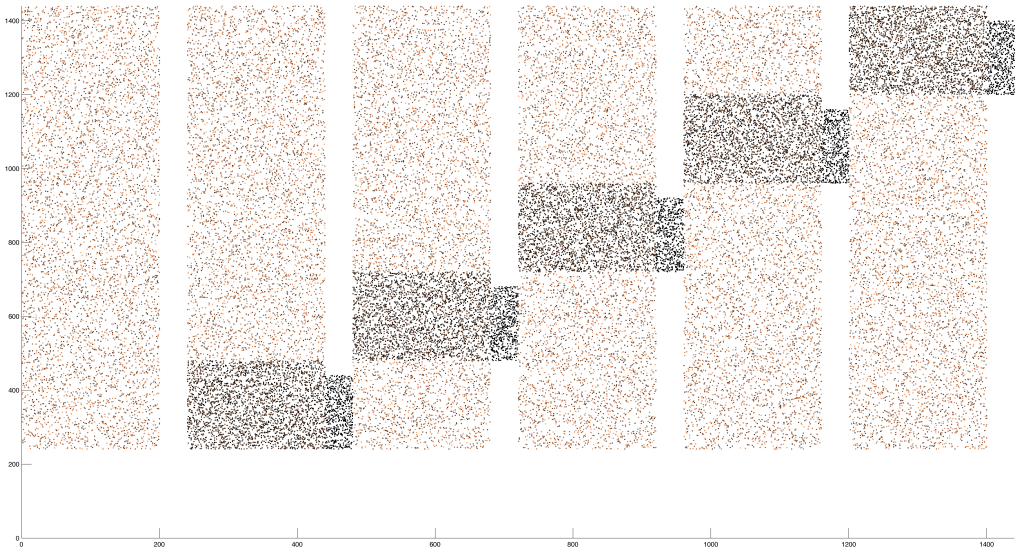

Figure 22: This plot portrays the network connectivity when Internal Delay Minimum is 1, Internal Delay Range is 8, Internal Connection Density is 14, External Delay Minimum is 7, External Delay Range is 18, and External Connection Density is 8. Axes refer to neuron index, where the horizontal axis corresponds to the presynaptic neuron, and the vertical axis corresponds to the postsynaptic neuron. Each point implies the existence of a projection, and the color of that point implies the extent of the action potential propagation delay (Ranging from black for a 1ms delay to copper for a 25ms delay.

architectures would vary in the firing patterns that emerged from different patterns of stimulation.

5.2 Analysis Metrics

Dynamics from each experiment were captured by recording the time and source of each spike produced by the network throughout the 1000 milliseconds of stimulation. A single experiment saw anywhere from 0 to around 350,000 recorded spikes, excluding the stimuli. To adequately comprehend the impact of changes across parameters, a suite of metrics were used to process data to facilitate its visualization.

The raw data is simply a list of times that each neuron spiked, shown rasterized in Figure 24 for a single run. Much can be seen in these spike rasters, but the phenomena observed can be difficult to quantify for comparison. Describing Figure 24, we see active spiking in all clusters that, beginning around 600ms, ramps up to a catastrophic explosion of activity which peaks near 700ms, the aftermath of which is complete silence as neurons return to equilibrium. This narrative of overall activity levels changing over time is well captured by a windowed histogram plot, shown for this run in Figure 25,

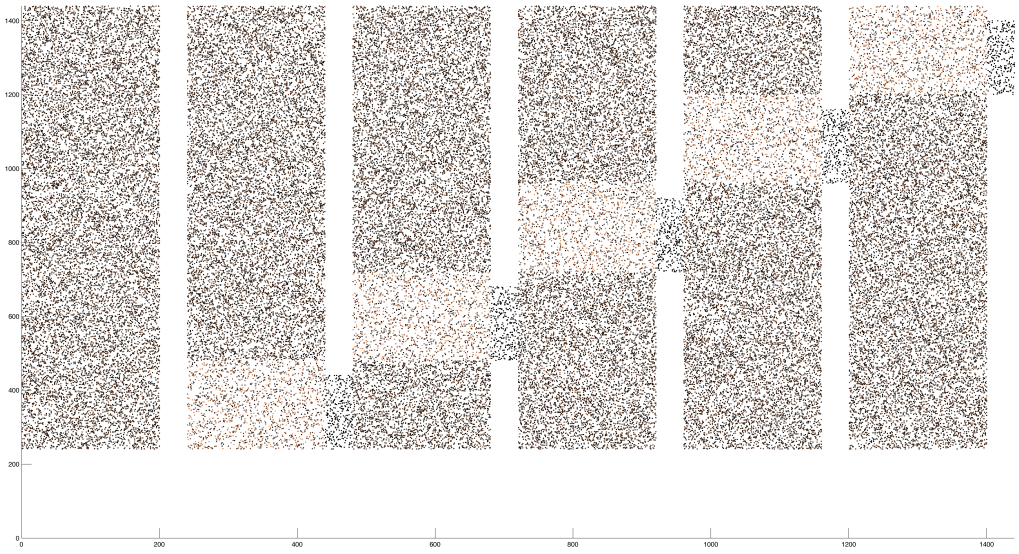

Figure 23: This plot portrays the network connectivity when Internal Delay Minimum is 1, Internal Delay Range is 18, Internal Connection Density is 8, External Delay Minimum is 1, External Delay Range is 8, and External Connection Density is 14.

which sums the number of spikes occurring within an 11ms window centered around each point in time. This sliding window approach smooths the resulting histogram, providing a clearer image of transitions in overall levels of activity. Notably, the explosion of activity around 700ms dominates this plot, dwarfing the visibility of all other spiking throughout the network. The similarity of dynamics across clusters motivates combining individual curves into a single, network-wide histogram. Statistics from the resulting histogram can then be used to compare different network results to assess the impact of parametric variation on the phenomena they describe.

The behavior of a network can be broadly characterized as being either supercritical, metastable, or subcritical. In the supercritical condition, like in Figure 24, networks exhibit explosive interactions which come to dominate dynamics to the extent that neurons over-stimulate to hyper-synchrony, massively extending their refractory periods and causing long windows of overall network silence between explosions despite continued stimulation. The patterns of these explosions of activity varied widely across parameter conditions, depending critically on network architecture. This contrasts with subcritical conditions, where stimulation was insufficient to cause significant downstream spiking. Between these extremes are metastable conditions, which respond enough from stimulation to cause

active spiking, but that spiking does not come to dominate network dynamics through catastrophic overstimulation.

It is of great interest to quantify and robustly identify the existence of these dynamics. This task is made considerably more difficult by a wide variety of explosive firing patterns. An explosion could appear to be more slowly rising and have a rounded peak, or it could have many rhythmic peaks, as in Figure 26. One distinct and powerful method utilized here involves employing the frequency spectra from the Fourier transform of the windowed histogram. Computing the magnitude of the coefficients using the discrete Fast Fourier Transform (FFT) for the histogram curve, like those in Figure 25 and Figure 27, yields the power level of the different frequencies of spiking, shown in Figure 28 and 29. Comparing the spectra of the histogram of the stimuli in Figure 28 to the spectra of other clusters in the network is striking: the clusters have very powerful low frequency spectra that dwarf the low frequency spectra of the stimuli. Computing the absolute value of the FFT on a 1000ms vector results in precisely 1000 real valued coefficients, symmetric around the 500hz, since all low frequency correlations can also be viewed as high frequencies (information is not known beyond 1000ms to disambiguate). This relationship implies that all explosive spiking dynamics will be pushed to the edges, as the sharpness of the histogram serves to amplify the power of the frequencies widened by its windowing. The spectra in Figure 29 are also massively powered in low frequencies, though the repetition of the explosions yield sharp peaks in higher, more moderate frequencies. To leverage this effect, these ultra-low frequencies can be summed and subtracted from those of the stimuli to identify the extent to which the network is over-stimulated, herein called the *histogram spectra metric*.

5.3 Parametric Analysis

How do parametric variations impact firing patterns? Since the impact of parameters are interdependent, and their resultant dynamics are so nonlinear, narratives for each parameter will need to be considered in terms of each other parameter for conclusions to become coherent regarding them. Some parameters, however, are far more defining of overall network activity than others. Let us consider these high impact parameters first.

The plot in Figure 30 provides the average number of spikes generated by a network, shown against the parameter condition index for all varied parameters, where each point averages all runs which used that parameter setting. This plot reveals which parameters can be changed to reliably increase or

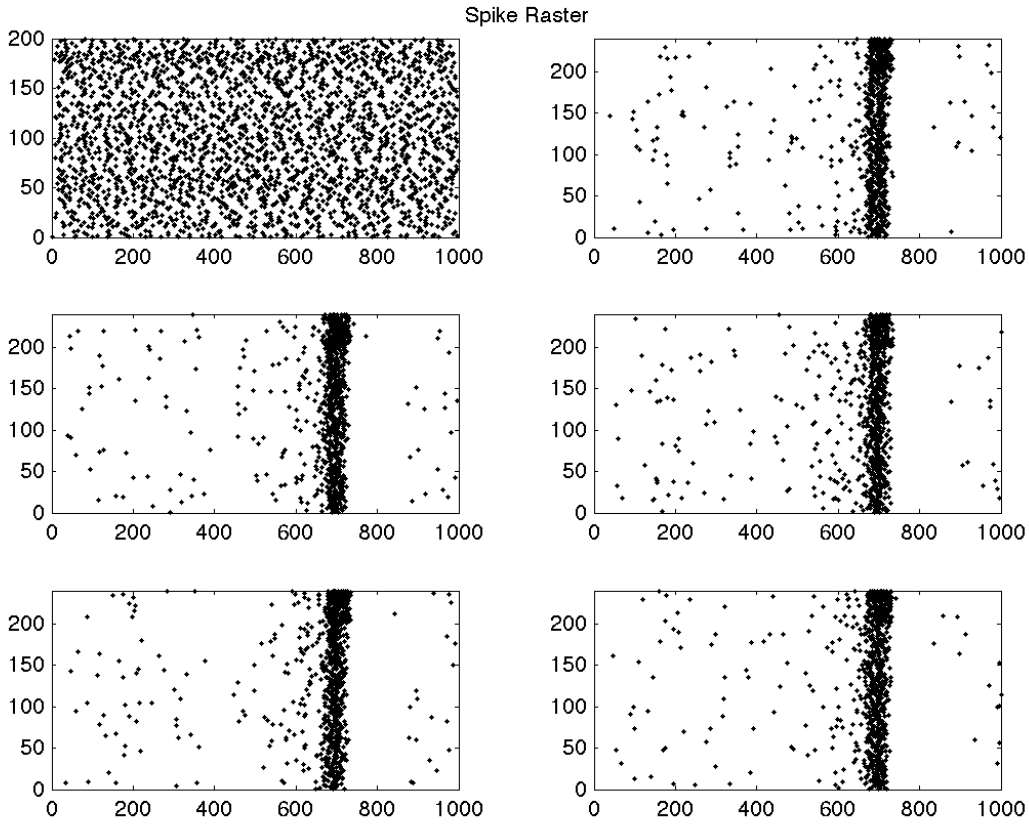

Figure 24: This is a raster plot of the spike data evoked from up to 25hz stimulation of a network where Internal Delay Minimum is 1, Internal Delay Range is 4, Internal Connection Density is 11, External Delay Minimum is 7, External Delay Range is 8, and External Connection Density is 11. Points correspond to existence of spikes, with neuron index plotted vertically and time plotted horizontally in milliseconds. The top-left plot is the source network, which stimulates the 5 interconnected subnetworks shown in the remaining plots. Neuron indexes 201–240 are fast spiking inhibitory neurons.

decrease the amount of spikes produced by the network. It is apparent that the strongest predictors of the number of spikes in the network are a high External Connection Density (ECD), a high Stimulus Frequency (SF), and low External Delay Range (EDR). It appears, however, that while Internal Delay Range (IDR), External Delay Mean (EDM), and Internal Connection Density (ICD) may influence the spike count of the network, their parametric variation does not necessarily guarantee the explosive conditions which dominate this plot. It is clear why External Connection Density should be the most influential factor here; External Connection Density determines the extent of all connections between clusters, *including the density of the projections from the driving stimulus.* In this regard, the ECD limits the bandwidth of network stimulation overall.

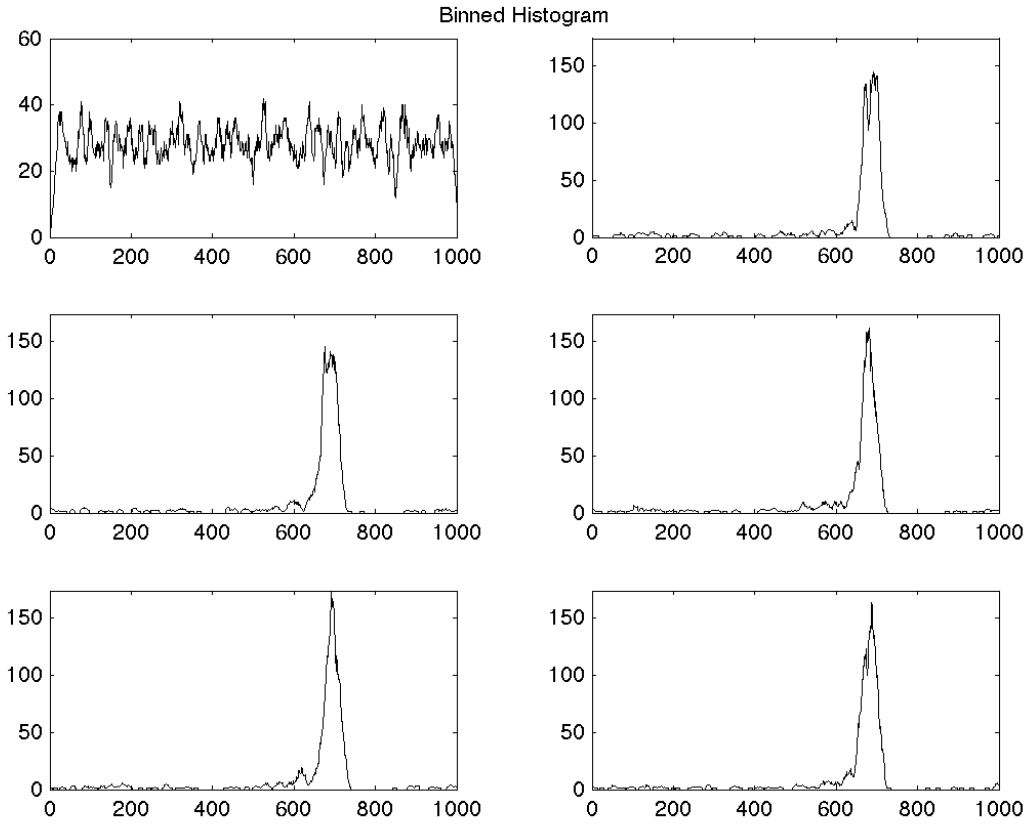

Figure 25: This is a histogram of the spike data portrayed in Figure 24. Each point sums all spiking activity within an 11 millisecond window centered on that point (+/- 5 milliseconds).

Notably, the top 40 networks that saw the highest amount of spikes in the experiment were all in cases with maximum External Connection Density and Stimulus Frequency, while having the lowest possible External Delay Range (which ensures that synchronized spikes are received at a synchronized time). Interestingly, these high spike count networks all had an External Delay Minimum of at least 7ms. The network which exhibited the most spikes is plotted in Figure 31. Seeing this spike raster clarifies why the highest spiking cases have External Delay Minimum of at least 7 ms—the existence of between cluster delays allows for the propagation of synchronous spike volleys that otherwise are silenced by locally acting fast spiking inhibitory neurons. Since all the delays are identical in the 0ms range case, the full volley is received all at once, ensuring a similarly hypersynchronous response. The delay is just long enough for each cluster to have quieted before receiving the new stimulation, though not long enough for ongoing stimulation to offset the rhythmic balance. Reducing the External

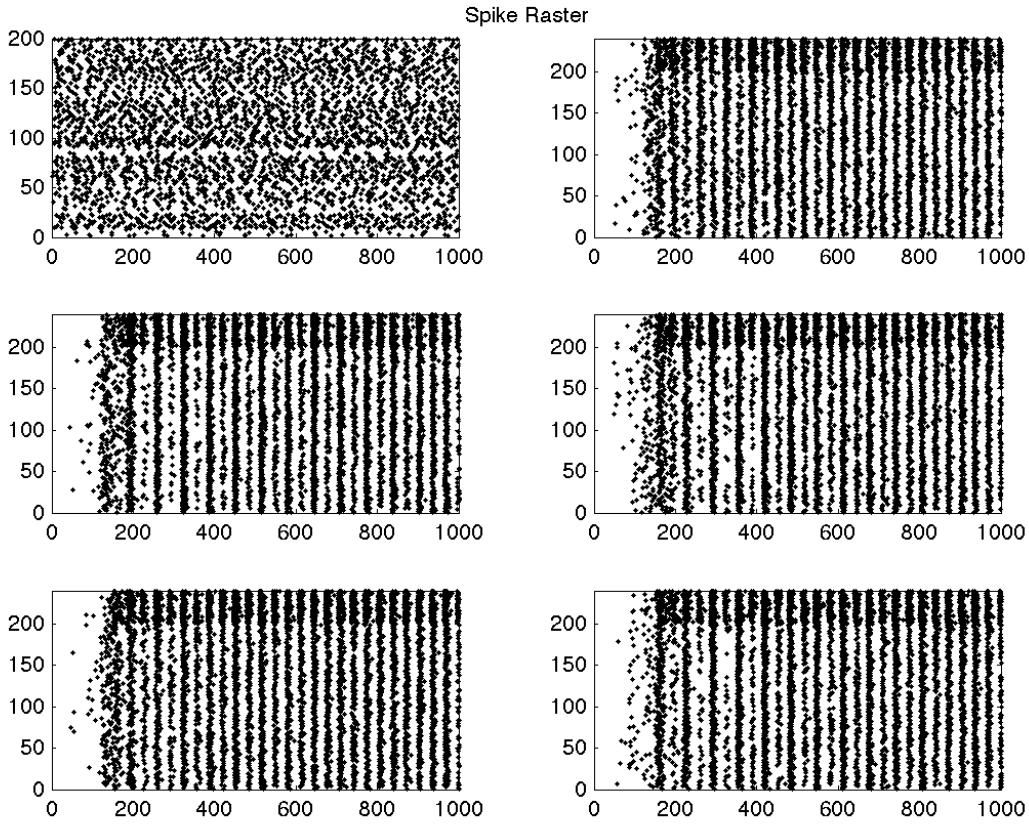

Figure 26: Spike data from up to 35hz stimulation of a network where Internal Delay Minimum is 1, Internal Delay Range is 8, Internal Connection Density is 14, External Delay Minimum is 25, External Delay Range is 8, and External Connection Density is 11.

Delay Minimum in this condition from 13ms to 7ms causes the initial wave of explosive synchrony to excessively reinforce itself, causing a massive 50ms explosion which repeats after a 200ms refractory period (down to a spike count of 211,286). If we increase the External Delay Range from 0ms to 4ms instead of reducing the External Delay Minimum, as portrayed in Figure 32, the impact is that synchronous waves are now received over 5 ms, causing the initial wave of explosive synchrony to cause explosions at a variety of times, resulting in a larger explosion which similarly reverberates until quieted from its earlier overstimulation (spike count of 170,933).

These examples reveal how large overall spike count is associated with activity which, while excessive, is just under the thresholds which would prevent further explosions. Another way of quantifying the network activity is to consider the maximum histogram value achieved throughout the experiment duration, plotted in Figure 33 against each parameter index. With this metric, each parameter is

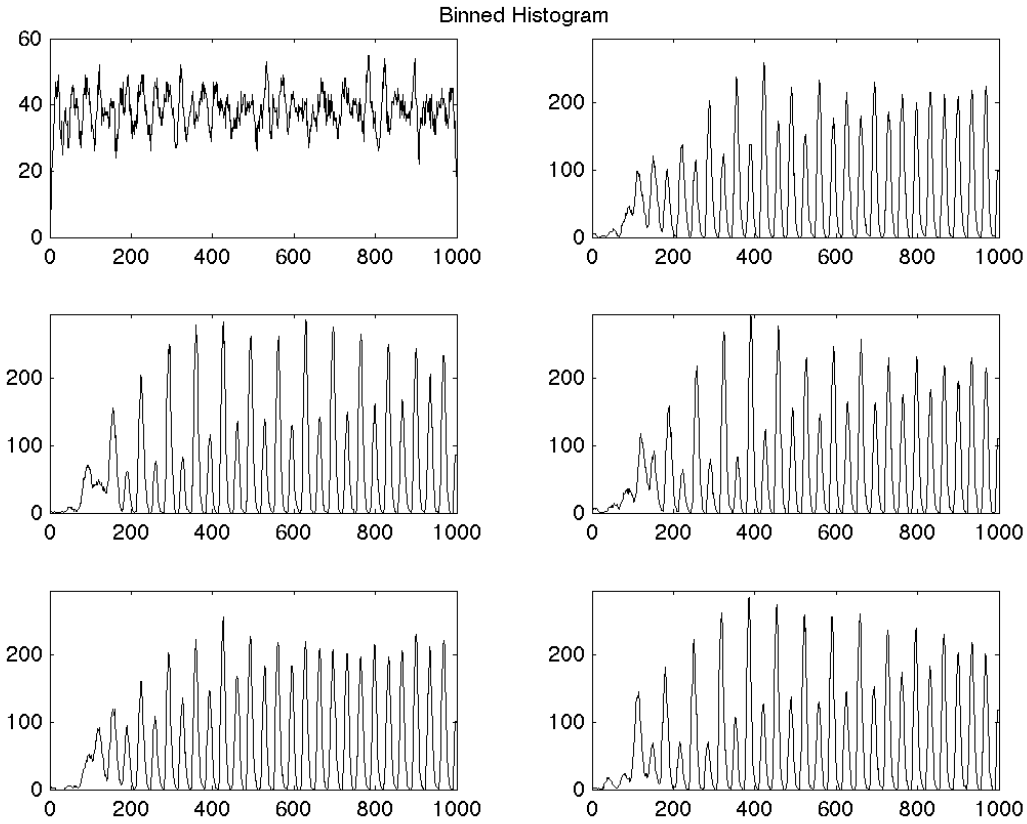

Figure 27: Histogram of the spiking data from Figure 26.

described in terms of its average contribution to explosive propensity. In Figure 33, in addition to the correlations expected from the spike count plots, we now see interesting relationships between the remainder of the parameters. Intuitively, increasing Internal Connection Density allows for larger explosive peaks, as its increase leads to an increased coupling within a cluster by increasing the likelihood that two locally evoked spikes will converge on a single neuron. Conversely, increasing the Internal Delay Range effectively *reduces* the size of peak explosions by reducing the likelihood that two locally evoked spikes will arrive at the same time, just like the External Delay Range parameter does for the network as a whole. Perhaps the most striking relationship noticeable here, as forecasted by Figure 31, is seen with the External Delay Minimum. When delays between clusters are allowed to be as short as delays within a cluster, then the network becomes as if it were one giant cluster—particularly when the External Delay Range is also small. Critically, these projections do not also connect to that cluster's inhibitory interneurons, leading to massively building levels of activity while partially undermining its

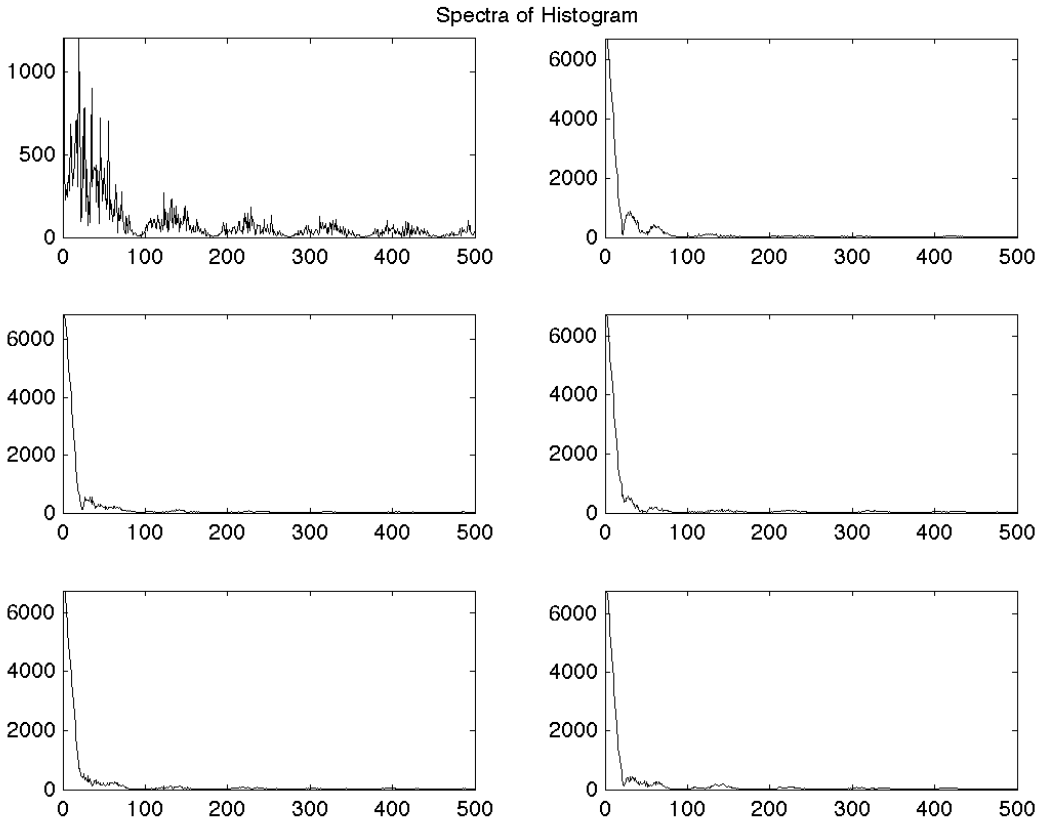

Figure 28: Frequency spectra of the windowed histogram in Figure 25. The horizontal axis corresponds to coefficients of the Fourier transform with its magnitude plotted vertically. This is used to compute the histogram spectra metric.

stability mechanisms. Increasing this the External Delay Minimum has the effect of providing time for the internal impact of a synchronous event to dissipate before spikes arrive from other clusters that were driven from a similar source. When the External Delay Range is 0, then the emergence of a hyper-synchronous spike volleys have a guaranteed synchrony upon receipt, causing reverberations like in Figure 31.

To illustrate the important distributional relationship between spike counts and maximum histogram values with respect to the External Delay Range, each parameter setting can be compared with a suite of scatter plots, like in Figure 34. From these plots, it becomes clear how the maximum histogram value decouples from the overall spike count in a network in a manner dependent on the External Delay Range. Since the maximum histogram value was computed from a binned histogram which counted spikes over 11ms over clusters of 240 neurons, the highest possible histogram value is

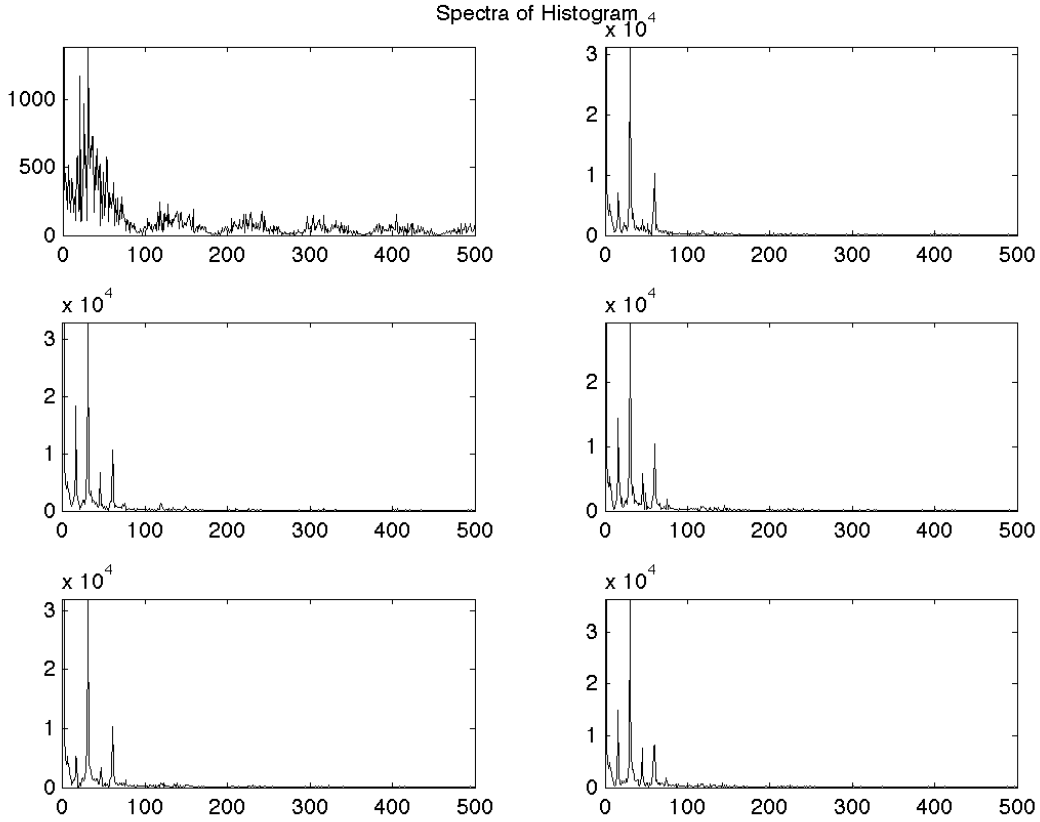

Figure 29: Frequency spectra of the windowed histogram in Figure 27. The horizontal axis corresponds to coefficients of the Fourier transform with its magnitude plotted vertically. This is used to compute the histogram spectra metric.

2,640, which forms the ceiling effect seen in most of the plots in Figure 34. Generally, it can be noted that the highest spike counts are runs which have lower histogram maxima, reinforcing the narrative implied by the results in Figure 31. In Figure 34, we see that as the External Delay Range gets larger, so does the correlation between the maximum histogram value and the spike count of the network (Figure 35).

All of the runs that achieved the cartoonish saturation levels required for a maximum histogram value of 2,640 were conditions with an External Delay Minimum of 1, which is something easily gleaned from Figure 36. The vertical striations present when External Delay Minimum is 7, 13, 19, and 25, are primarily from conditions when External Delay Range is 0, and are a result of stereotyped oscillations that are present in those conditions.

The most consistent metric to identify the existence of explosions in the data was the histogram

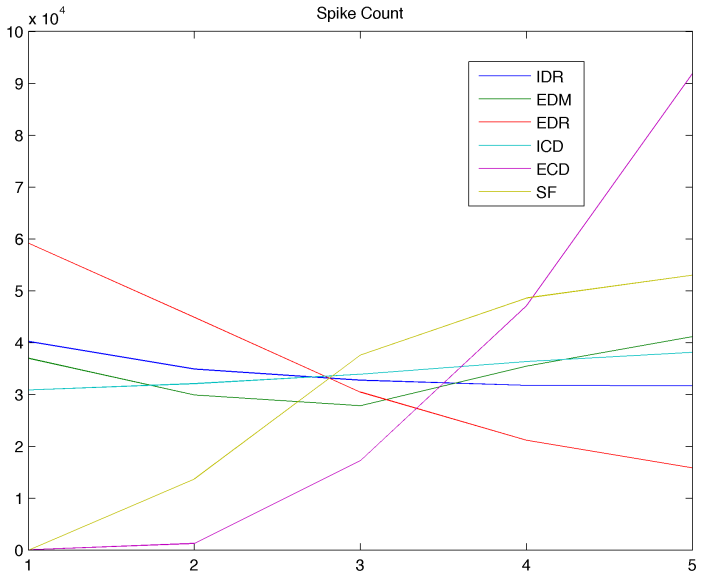

Figure 30: Average number of spikes generated for each parameter condition.

spectra metric. Figure 37 shows how changes in the External Delay Minimum parameter adjust the relationship between spike count and the histogram spectra. Consistent with previous analysis, conditions are most explosive when the External Delay Minimum is 1. These are the explosions that are most saturated, leading to large gaps where neurons cannot be further stimulated. When the External Delay Minimum becomes larger, explosions take longer to propagate, allowing the network to become more stable between volleys. Figure 37 reveals how the histogram spectra metric separates these low frequency explosions from smaller, more frequent explosions, identifying differences in firing regime dependent on the External Delay Minimum parameter. The pattern evident in Figure 38 shows how External Connection Density exaggerates the External Delay Minimum parameter as the number of connections per neuron increases by either increasing the rate of explosive oscillation, or by increasing the magnitude of the explosion, in a way that depends on the firing regime. The pattern of firing exhibited by a network appears to be best described by the way its parameters change the likelihood that two spikes will be coincident upon receipt, determining the structure of explosive propagation. Since External Connection Density, Internal Connection Density, and Stimulus Frequency only change the *number* of spikes that can propagate in the network, not *how* those spikes are received, then these parameters ostensibly *amplify* the number of spikes flowing through a particular regime *without determining the structure* through which those spikes flow. This amplification does, however, serve to

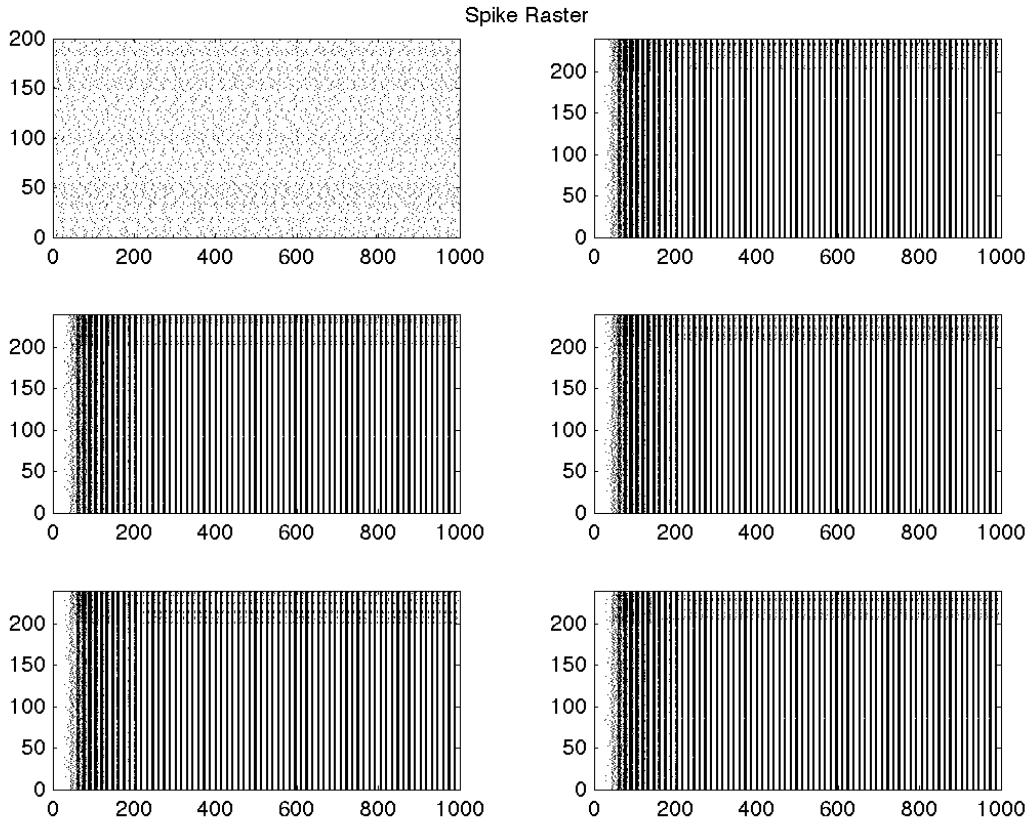

Figure 31: Spike data from up to 45hz poissonian stimulation of a network where Internal Delay Minimum is 1, Internal Delay Range is 0, Internal Connection Density is 17, External Delay Minimum is 13, External Delay Range is 0, and External Connection Density is 17. It generated the most spikes in its data set, with a total of 349,555.

shift the boundaries within which the network becomes either subcritical (understimulated), responsive (moderately stimulated), or supercritical (overstimulated).

5.4 Implications

How do the firing patterns emerging from recurrently connected and clustered network architectures respond to systematic changes in connectivity? Simulation data from an extensive parametric analysis of connectivity and delay patterning suggest that a lack of variability in external conduction delays, combined with short external delay means, produce networks that are massively excitable. These networks respond extremely to stimulation, resulting in distinct firing patterns that wash out information as they become stereotyped. Conversely, networks with longer external delay means, combined

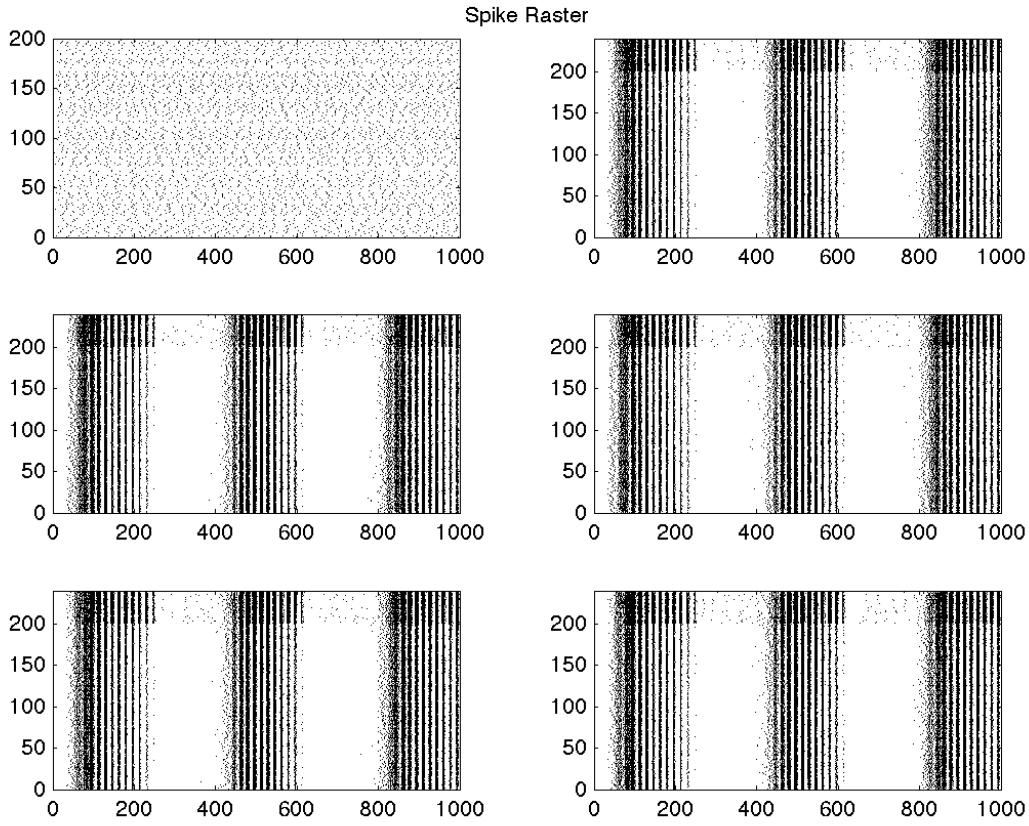

Figure 32: Spike data from up to 45hz poissonian stimulation of a network where Internal Delay Minimum is 1, Internal Delay Range is 0, Internal Connection Density is 17, External Delay Minimum is 13, External Delay Range is 4, and External Connection Density is 17.

with a wider range in those external delays, results in a network that can withstand more stimulation while having more connections per neuron. This result suggests that variability in action potential propagation delays in cortex may enable a greater density of connectivity between seemingly disparate subregions with less threat to the stability of the larger network.

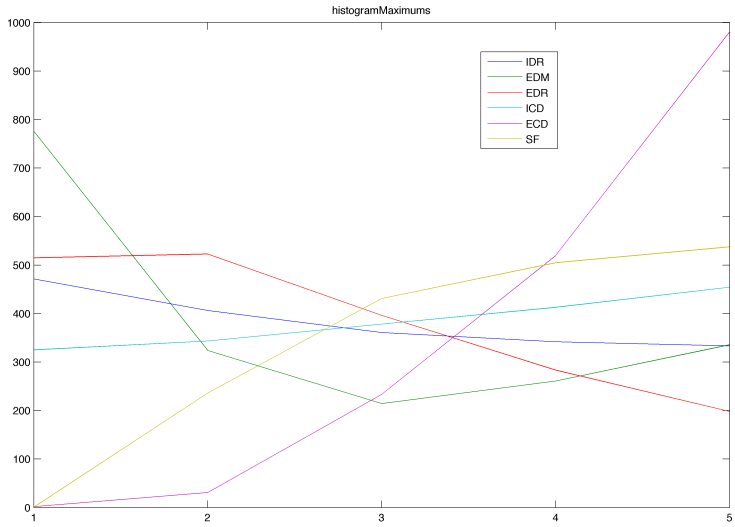

Figure 33: Average maximum histogram value of spikes generated for each parameter condition.

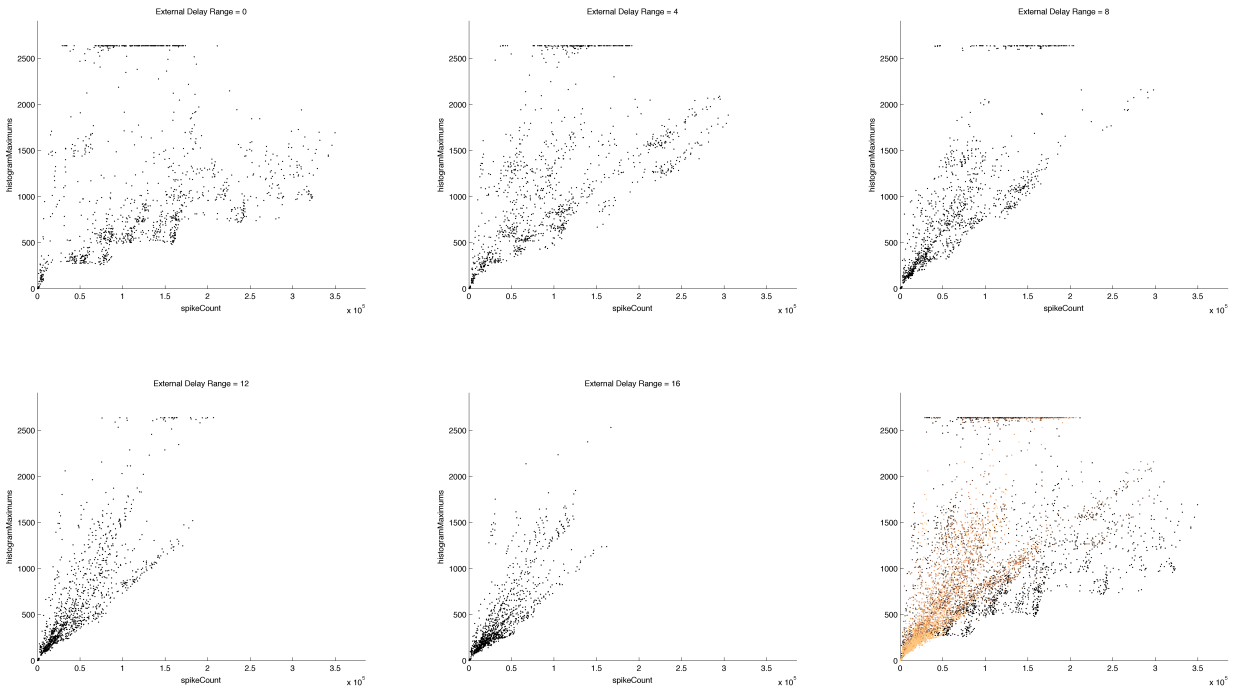

Figure 34: In these scatter plots, each point corresponds to a single run with a particular External Delay Range, where the vertical axis corresponds to the run's maximum histogram value, and the horizontal axis corresponds to the run's total spike count. Plots are correspond to External Delay Range values 0, 4, 8, 12, and 16 (from top-left to bottom-middle). The final plot shows all 5 parameter values plotted together, with low ranges in black, and high ranges in copper.

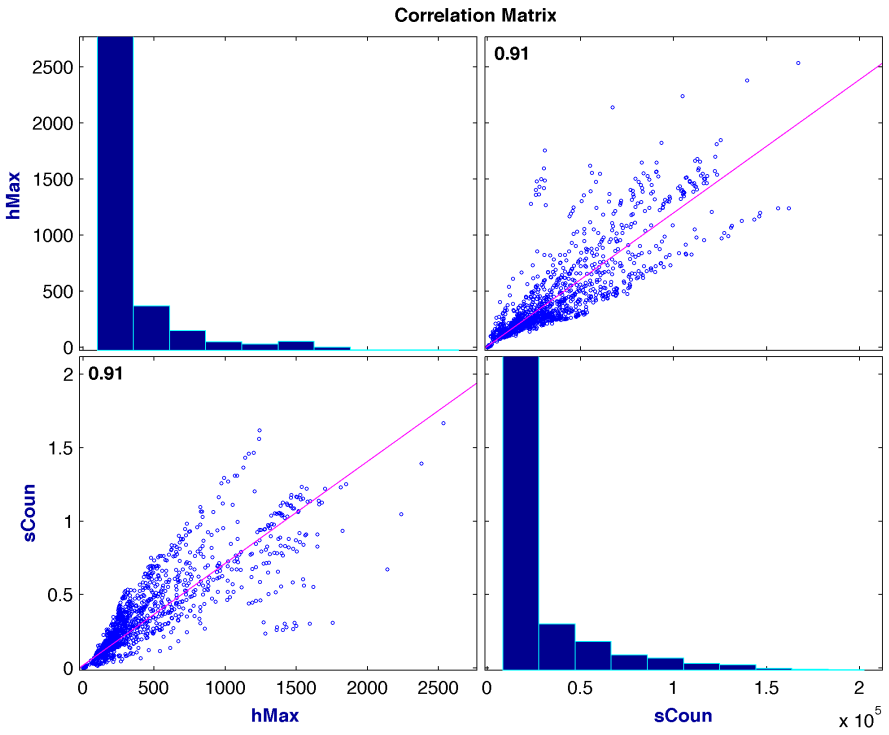

Figure 35: This is a correlation plot between the maximum histogram values and the spike counts when External Delay Range is 16, where plots on the main diagonal are histograms of the data, and the correlation plots are plotted with a line with slope 0.9092 determined from least-square regression. When External Delay Range is 0, the slope becomes 0.6668.

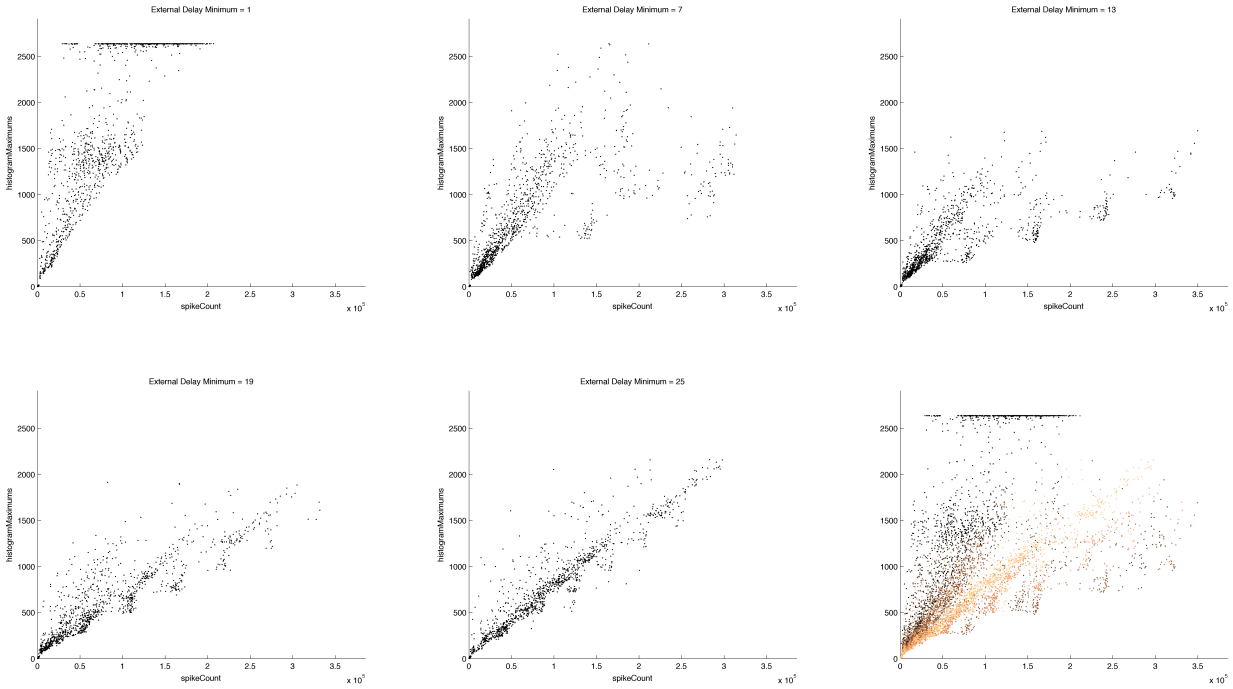

Figure 36: Scatter plots of maximum histogram values against spike counts, separating runs by their External Delay Minimum parameter. Plots correspond to values 1, 7, 13, 19, and 25 (from top-left to bottom-middle). The final plot shows all 5 parameter values plotted together, with low ranges in black, and high ranges in copper.

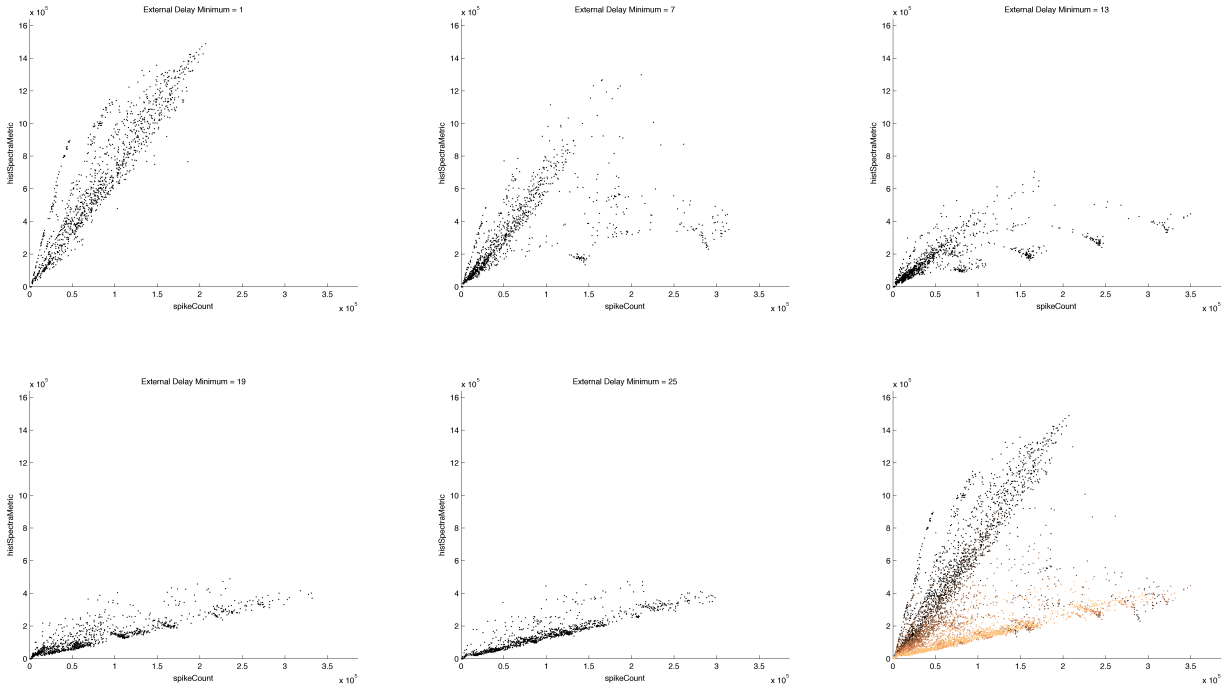

Figure 37: Scatter plots of spike counts against the histogram spectra metric, separating runs by their External Delay Minimum parameter. Plots correspond to values 1, 7, 13, 19, and 25 (from top-left to bottom-middle). The final plot shows all 5 parameter values plotted together, with low ranges in black, and high ranges in copper.

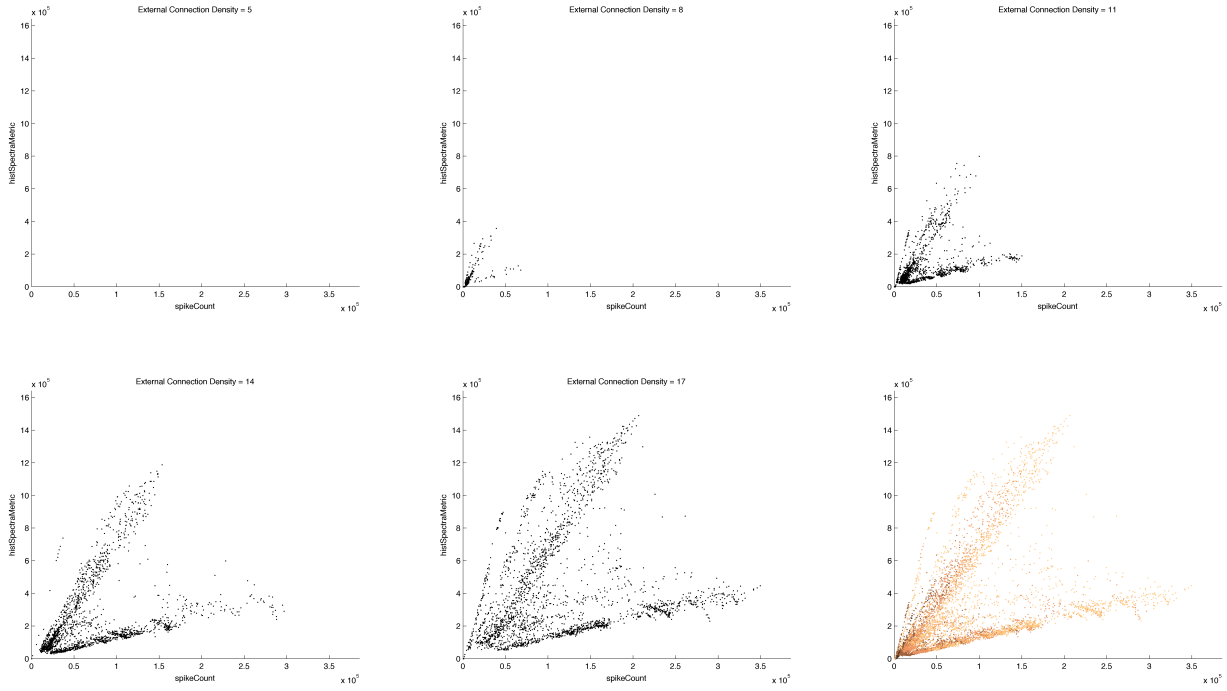

Figure 38: Scatter plots of spike counts against the histogram spectra metric, separating runs by their External Connection Density parameter. Plots correspond to values 5, 8, 11, 14, and 17 (from top-left to bottom-middle). The final plot shows all 5 parameter values plotted together, with low ranges in black, and high ranges in copper.

6 How the Neural Code Depends on Network Topology

In Section 5, a parametric analysis of a clustered network architecture revealed the impact of varying network connectivity patterns and delays on emergent firing patterns and network excitability. Inferences, however, were focused on identifying how network parameters contributed to the most explosive dynamics. Ultimately, these explosive contributions were best described by the way that parameters shifted the likelihood of coincidence upon receipt of spikes in the network. This was achieved either by increasing the likelihood that spikes will be sent, as a result of increasing the stimulation frequency or the number of projections, or by increasing the synchrony of spike receipt, as a result of reducing propagation delay or its variability. In a supercritical network, however, it's unclear how information may propagate through the saturation of hypersynchronicity. Once a network explodes, the stereotyped nature of that explosion obscures *how* that network was originally stimulated, indicating simply that it achieved sufficient stimulation to explode.

The success of a neural code depends on its ability to distinctly propagate. If we assume that information is best seen as a firing-rate based population code, then hypersynchronous explosions force a single rate dynamic to the affected population, yielding a convergent rate code dependent only on the extent of original stimulation. If we assume information is seen as a heteroclinic chain of spikes, as in PNGs, then hypersynchronous explosions force a single heteroclinic chain, yielding a convergent channel that is not reflective of the distinct way in which it was triggered. In both cases, supercriticality undermines the ability for the network to respond to information in a way distinct to that information, resulting in catastrophic equivalency. Similarly, subcriticality leads to understimulation, equivocating all but the most extreme stimulations through quiesence. Thus, neural coding is expected to be most effective *between* these two dynamics, where the network is active enough to propagate distinctly, but not so active as to wash out its significance.

6.1 Identifying Active Networks

To assess the viability of neural coding in different clustered networks, we must identify the networks in which propagation is at all possible, herein called *active networks*. This requires the ability to exclude both supercritical and subcritical networks from analysis. In Section 5.2, it was demonstrated how the histogram spectra can be used to identify the existence of explosions in firing patterns. A metric made from these spectra was used in Section 5.3 to characterize the impact of parameters on explosive

excitability. Here, the histogram spectra metric will be used to determine a diagnostic threshold to identify the extent to which a particular clustered network exhibits supercritical dynamics.

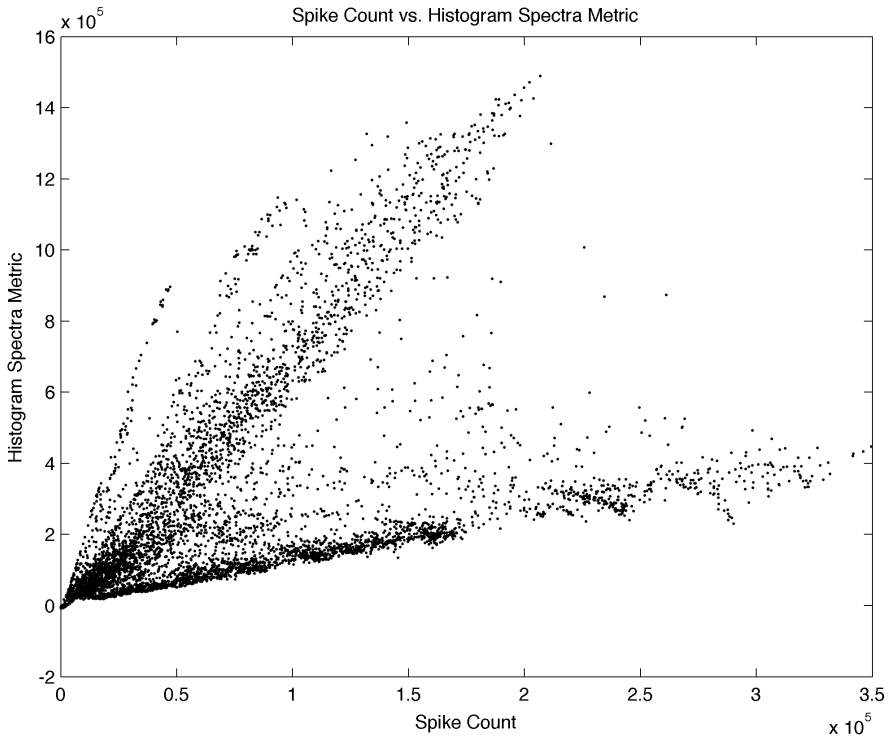

Figure 39: Spike count plotted against histogram spectra metric for each run.

While there is interesting structure present in Figure 39, nearly all data points that can be distinguished are supercritical. This is because the scale of these explosive runs is much larger, on the order of 10^5 spikes. However, a little over 61% of all runs have a histogram spectra metric of *less than 0*, revealing how the distribution of most runs will only be visible by zooming in on the bottom left corner of the plot. In Figure 40, the cluster below 0 on the vertical axis can be clearly seen. Furthermore, the distribution is fairly isolated—it has a large gap between it and when the major distribution of explosive runs begins.

To determine a value for the histogram spectra metric that is diagnostic of supercriticality, the lowest value of the histogram spectra metric for which the network is explosive needs to be found. All networks with a histogram spectra metric less than 0 are ones that have network histogram spectra that is less than the histogram spectra of the stimulus drive, since the metric is the difference of one from the other. Thus, if it has a histogram spectra metric of less than 0, it is known that it is at

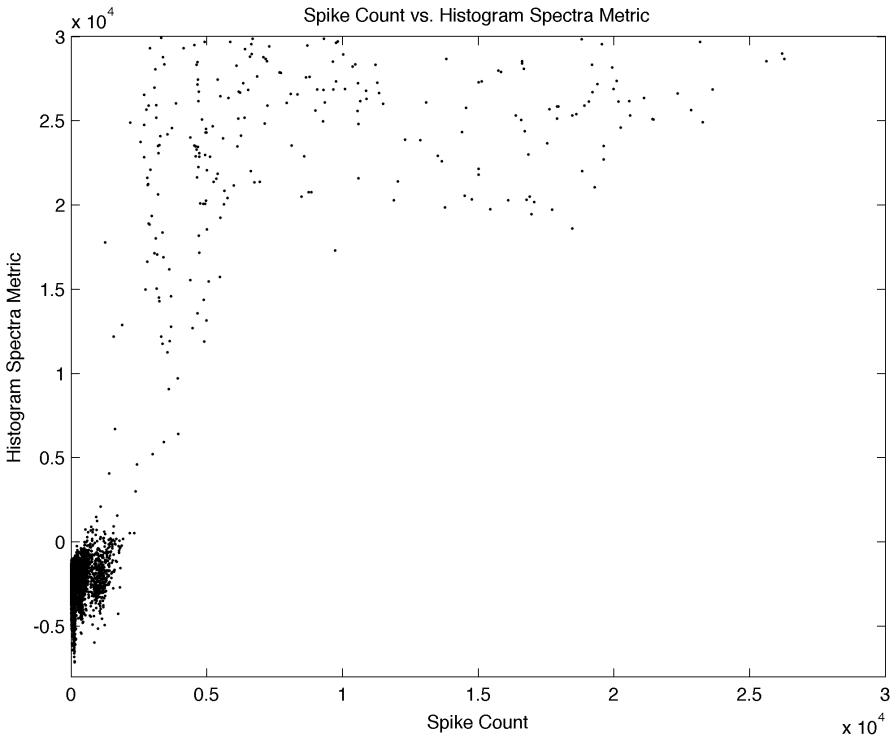

Figure 40: Spike count plotted against histogram spectra metric for each run, zoomed in with the horizontal axis restricted to being from 0 to 30,000, and the vertical axis restricted from -8,000 to 30,000.

least less explosive than its stimulus. In Figure 41, we can only see the 41 runs that lie between 0 and 5,000 with respect to their histogram spectra metric. 37 of the runs have a histogram spectra metric between 0 and 2,000, plotted in Figure 42—there is no evidence of explosions in these runs. All 4 runs between 2,000 and 5,000 of the histogram spectra metric exhibit dynamics which ramp up suddenly, however they do so far more gently than the explosive dynamics discussed so far. In Figure 43, a binned histogram reveals how spikes begin to ramp up in the last 100 milliseconds of the simulation. If continued, this run would peak, and then be followed by a period of quiescence, resembling what is observed in the run with histogram spectra metric value 3,006 (plotted in Figure 44). Plotting all runs with histogram spectra value 5,000 through 10,000 clarifies that, indeed, all runs with histogram spectra value 2,000 or greater show signs of at least gentle explosive dynamics followed by network silence. This suggests that using the upper threshold of 2,000 for a network's histogram spectra metric will exclude explosive dynamics, and include all runs which exhibit robust responses.

The histogram of the spike counts of all runs with a histogram spectra value between 0 and 2000

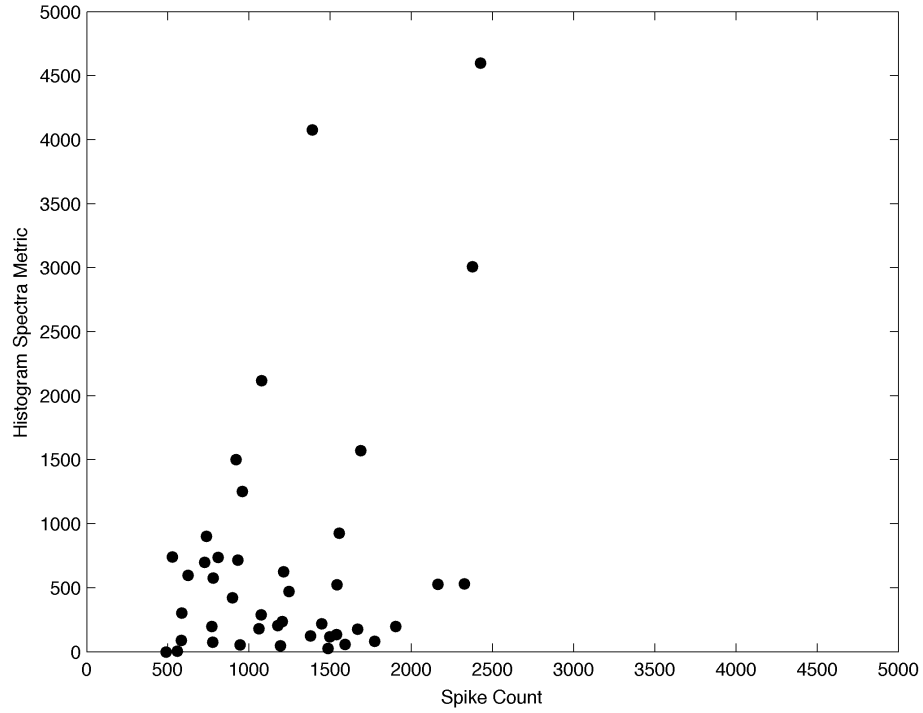

Figure 41: Spike count plotted against histogram spectra metric for each run, zoomed in with the horizontal axis restricted to being from 0 to 5,000, and the vertical axis restricted from 0 to 5,000.

are plotted in Figure 45. To remove subcritical networks, a spike count filter can be applied. Limiting runs to at least 20 spikes shortens the list from 9676 to 4729 runs. Further increasing it to 200 reduces the list to just 1,637 different runs, with the distribution of spike counts shown in Figure 46. Figure 47 shows the included subset in black, and the excluded subset in red on a plot of histogram maximums against spike count.

6.2 Describing Active Networks

Now that we have identified the active cluster networks, we can describe the network factors which prevent their explosion, and describe their dynamics. In Table 3, each column corresponds to a count of active networks with a different parameter setting (with value implied in Table 2), providing insight into which parameters are most constrained by the filter. Particularly, the parameters External Connection Density and Stimulus Frequency are counter to each other, where if one is large, the other must be small to facilitate significant dynamics. As they get closer together, other parameters shift down to "compensate," like Internal Connection Density or Internal Delay Range. Increasing the spread of

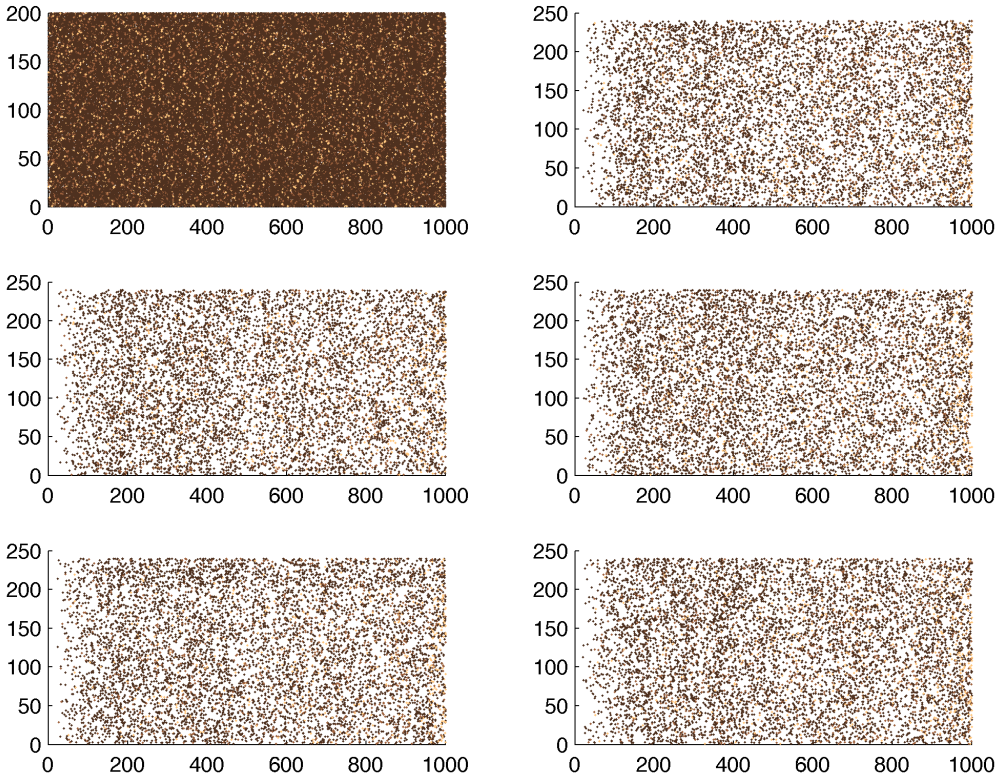

Figure 42: All spike data (37 runs) with histogram spectra metric between 0 and 2,000, with External Connection Density parameter values highlighted from black to copper. There does not appear to be any explosions within the data.

delays has the impact of allowing activity to dissipate, since by having a larger range of delays leads to increased sparsity from the reduced synchrony of spike receipt. These factors work together to enable a fully recurrent spiking network that responds to constant stimulation without leading to catastrophy. In Table 4, we see a table similar to Table 2, but now only networks that were metastable under 45hz stimulation are counted. With such large tonic stimulation, clustered networks are more robust due to less internal resonance from decreased internal connectivity, and have higher External Delay Ranges and External Delay Minimums.

6.3 Stability of Firing Rate

The identification of active networks allows for an analysis of the stability of the rate coded persistant stimulation of those networks. How can a network be said to be unstable or stable with respect to a rate

	1	2	3	4	5
ICD	5	8	11	14	17
IDM	1				
IDR	0	4	8	12	16
ECD	5	8	11	14	17
EDM	1	7	13	19	25
EDR	0	4	8	12	16
SF λ	5	15	25	35	45

Table 2: All tested parameter settings: Internal Connection Density, Internal Delay Minimum, Internal Delay Range, External Connection Density, External Delay Minimum, External Delay Range, and Poissonian Stimulus Frequency.

	1	2	3	4	5
ICD	391	381	333	284	248
IDM	1637				
IDR	318	321	327	337	334
ECD	0	974	510	81	72
EDM	281	328	336	342	350
EDR	282	313	347	337	358
SF λ	0	153	510	619	355

Table 3: Distribution of parameter settings that construct an active network. Corresponds to the number of active networks with a particular parameter setting (out of 15,625).

	1	2	3	4	5
ICD	119	109	69	38	20
IDM	355				
IDR	63	74	71	70	77
ECD	0	355	0	0	0
EDM	49	69	74	79	84
EDR	52	68	79	78	78
SF λ	0	0	0	0	355

Table 4: Distribution of parameter settings that construct an active network when stimulated at 45hz.

	1	2	3	4	5
ICD	29	29	35	24	27
IDM	144				
IDR	26	24	29	37	28
ECD	0	0	0	75	69
EDM	18	31	30	37	28
EDR	22	22	32	28	40
SF λ	0	144	0	0	0

Table 5: Distribution of parameter settings that construct an active network when stimulated at 25hz.

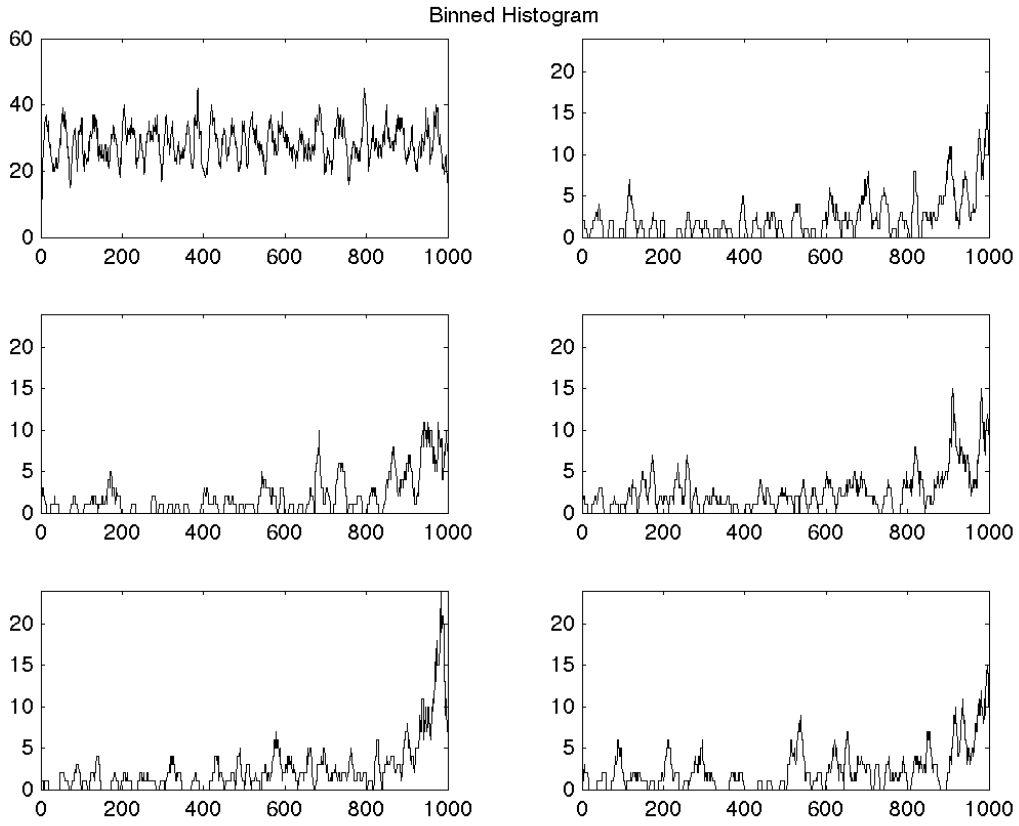

Figure 43: This is the binned histogram for the run with a histogram spectra metric value of 2,116. In the last 100 milliseconds, activity can be seen ramping up to what will likely become an explosion if the simulation were to be run longer. Parameters: Internal Delay Minimum 1, Internal Delay Range 12, External Delay Minimum 19, External Delay Range 12, Internal Connection Density 17, External Connection Density 11, and a Stimulation Frequency of 25hz.

coding information scheme? This question can be addressed by considering the stability of interspike intervals for particular networks. A rate-based population code depends on having sufficient activity to drive the dynamic of interest. Let us first consider the most active networks for examination.

In Figure 48, the mean interspike intervals for each neuron in an active run are plotted with their standard deviations. The excitatory neurons for each cluster have been sorted by their interspike interval, with a maximum interspike interval threshold of 250ms. Neurons that did not fire at least 3 times within the 1,000 milliseconds were ignored. The stimulus can be see in the top-left, which reveals the consistency and stability of the network's stimulation. Recognize that these are in terms of interspike intervals, so large standard deviations have an even larger impact when they are used to estimate the expected number of spikes per second (150 milliseconds of interspike interval variability

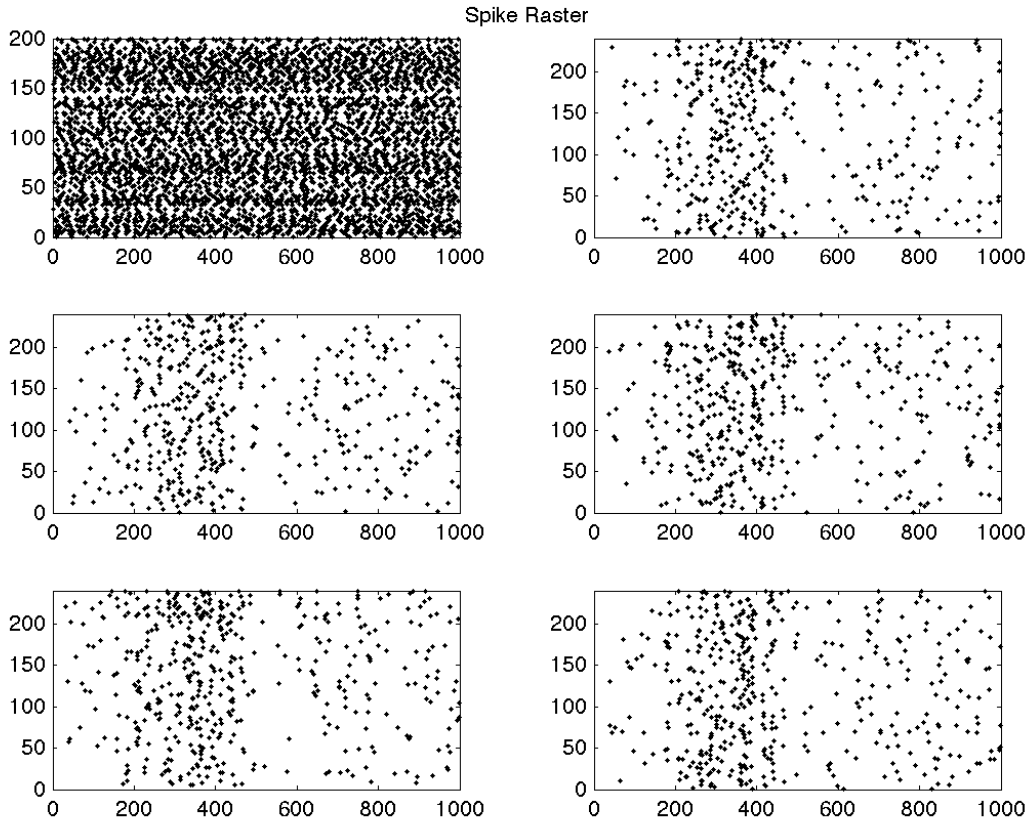

Figure 44: Spike raster of the run with a histogram spectra metric value of 3,006. A large bump in activity can be observed, followed by a period of silence. Parameters: Internal Delay Minimum 1, Internal Delay Range 8, Internal Connection Density 11, External Delay Minimum 19, External Delay Range 0, External Connection Density 8, and a Stimulation Frequency 45hz.

on 100 milliseconds becomes 6 Hz variability on a 10 Hz signal). Comparably, the network portrayed in 49, which represents the most active non-explosive network architecture, also shows poor promise for propagating rate-related information, where its mean standard deviation is 29.1559.

In Figure 50, the mean standard deviation of interspike interval (red) is plotted along with its standard deviation (black), and below its mean interspike interval (blue). This plot reveals a common pattern among clustered networks: low standard deviation of interspike interval is is often indicative of a low firing rate overall. This is generally illusory, however, as it is an artifact of having just enough spikes to statistically register an interspike interval, though without producing enough spikes for the interval to actually be resampled.

After examining the distributions of mean interspike intervals present in Figure 51, it is clear that

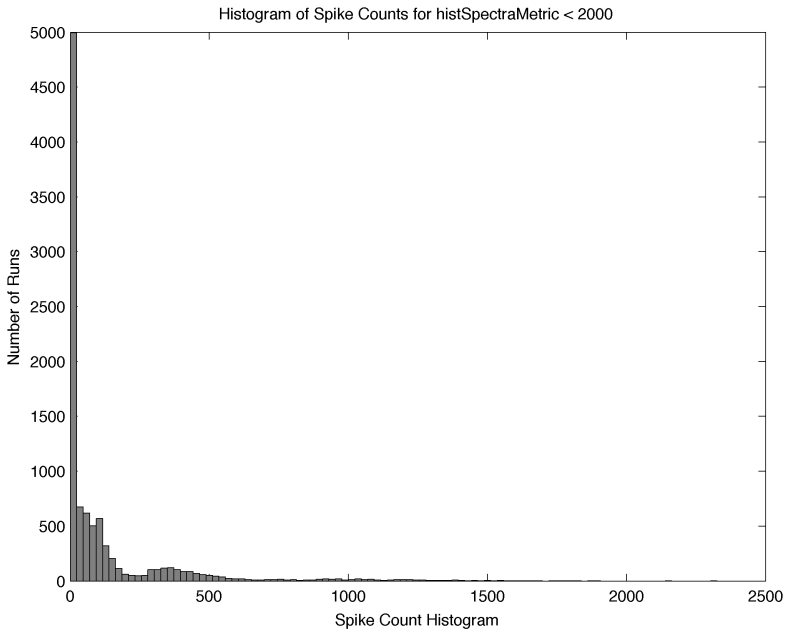

Figure 45: Histogram of Spike Counts for all runs with a histogram spectra metric value < 2000.

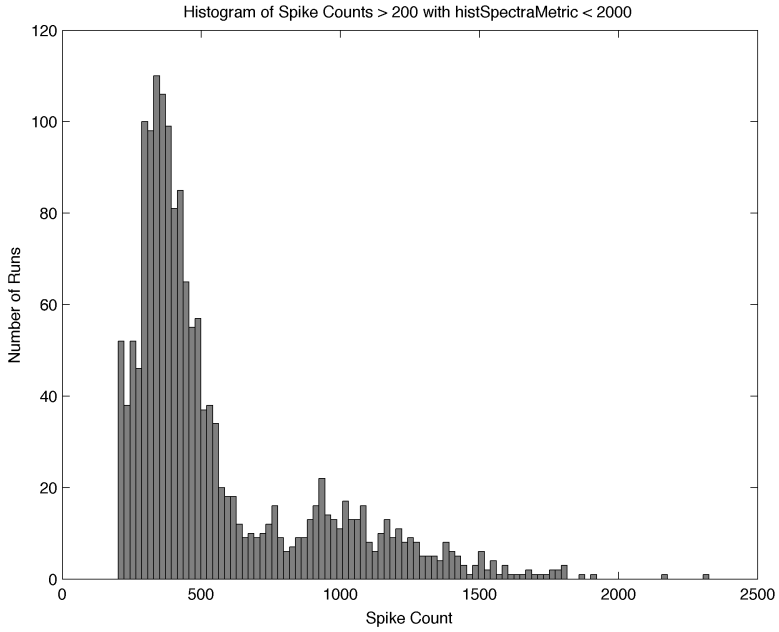

Figure 46: Histogram of Spike Counts > 200 for all runs with a histogram spectra metric value < 2000.

blind averaging across neuron interspike intervals will result in data skewed by neurons that fire within a short period of time, but never again firing, resulting in the appearance of a short interspike interval with low standard deviation. Increasing the required number of interspike intervals does not solve

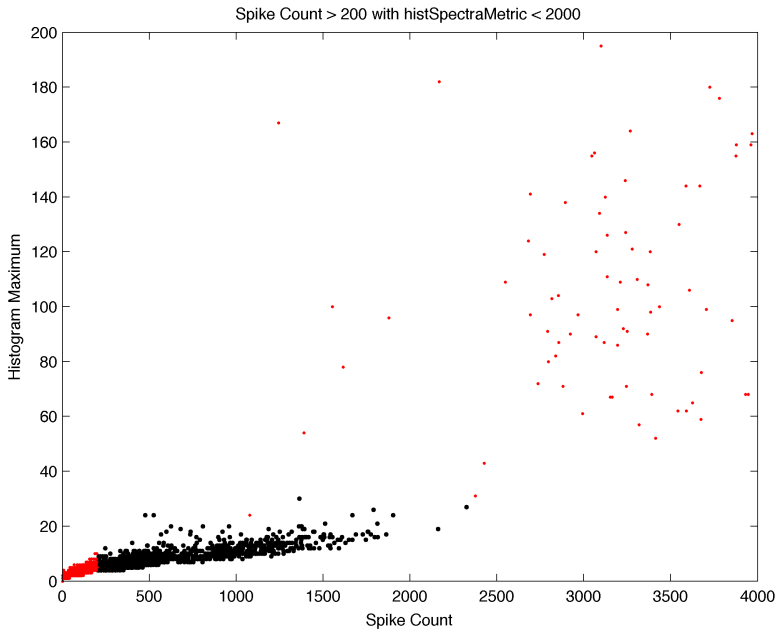

Figure 47: Scatter plot of spike counts against histogram maximums, where all runs with Spike Counts > 200 and histogram spectra metric value < 2000 plotted in black, and all other runs plotted in red.

the problem, however. Increasing the required number of interspike intervals per neuron decreases the likelihood that a set of interspike intervals will be recorded at all, and does not change the likelihood that a set of neurons may stop firing all together.

To avoid these degenerate cases, a method was utilized which compares the amount of spikes produced over a time window with the number of spikes implied by that neuron maintaining its rate code for the 1000 millisecond duration. Neurons that suggest a mean interspike interval of 20 milliseconds for a 1000 millisecond duration should produce some portion of 50 spikes total spikes. The *Rate Code Tolerance* (RCT) parameter shifts this proportion of spikes required by $\frac{RCT}{100}$. In Figure 52, different RCT values are tested for how they categorize ISI stability of runs (and hence networks). The horizontal axis corresponds to different runs, and is sorted based on rank inclusion of interspike intervals. Colors correspond to the lowest level of tolerance required to include the interspike interval within the filter, computed based on the Rate Code Tolerance listed at the top of each plot. This reveals that having RCT values more than 50 results in a substantial number of excluded runs that have larger numbers of interspike intervals, which can be seen by the vertical bars on the left folding into the included distributions as the considered RCT values decrease. In the bottom right of Figure 52, the number of networks included in the analysis is plotted against different RCT values.

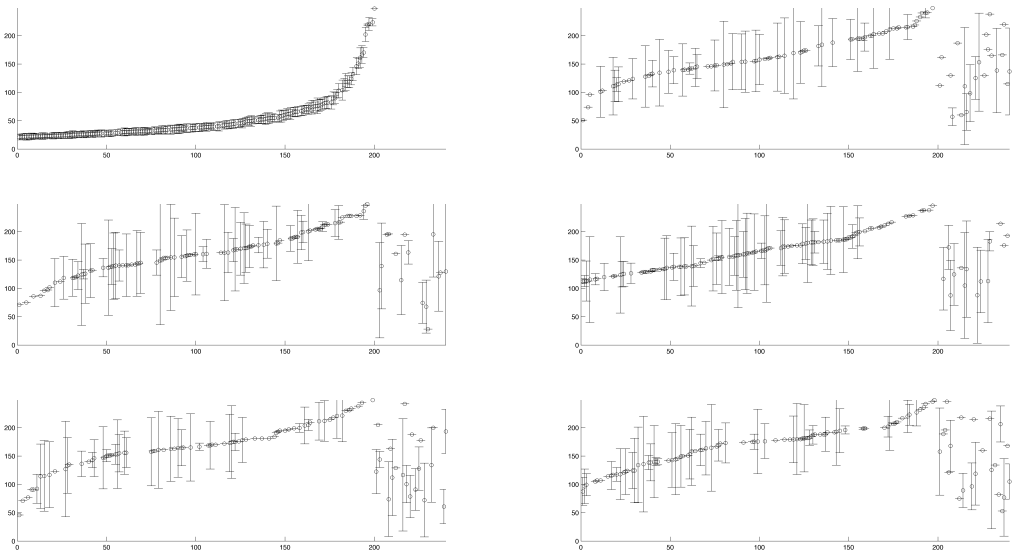

Figure 48: This run contains 1,863 spikes, and is one of the most active runs. This plot portrays the mean interspike interval for each neuron in the network, with bars portraying standard deviation. The excitatory neurons for each cluster have been sorted by their interspike interval, with a maximum interspike interval threshold of 250ms. Inhibitory neurons (201–240) are separated for each cluster. Mean standard deviation of 18.8821, itself with a standard deviation of 24.4567.

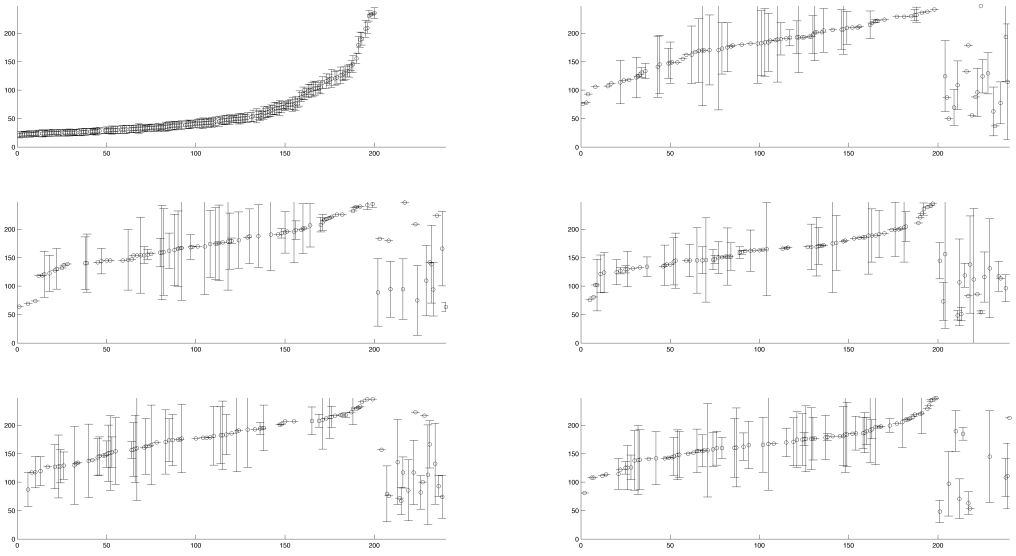

Figure 49: This run contains 2,325 spikes, and is the most active run that is not excluded. This plot shows the mean interspike interval for each neuron with its standard deviation.

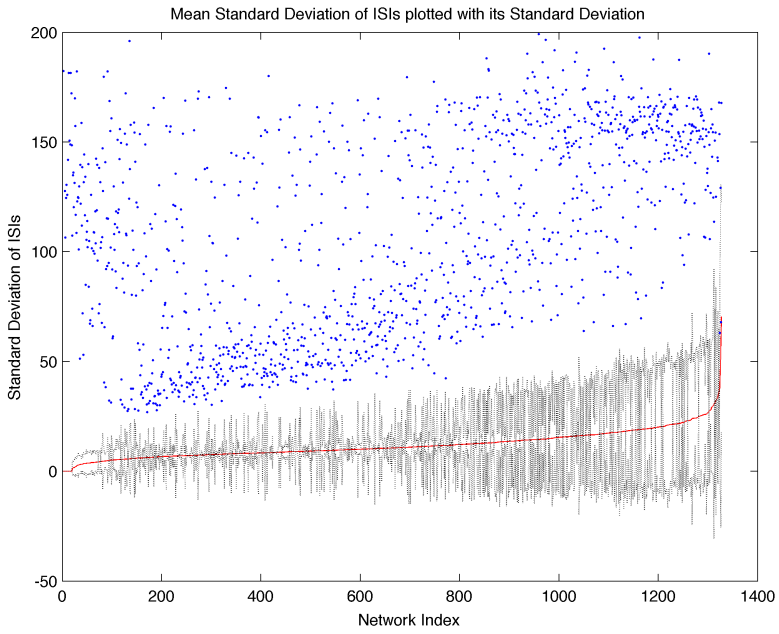

Figure 50: This shows the mean standard deviation of interspike intervals for all active networks in red, plotted with standard deviations surrounding it. The networks are sorted by mean standard deviation of interspike intervals. Each network's mean interspike interval is plotted as a blue dot above the curve.

Around an RCT of 30, there is a steep jump in the number of unique runs included, identifying a value where the number of interspike intervals and networks jumps up dramatically if continued. Using RCT of 30, we can more comfortably aggregate interspike interval statistics.

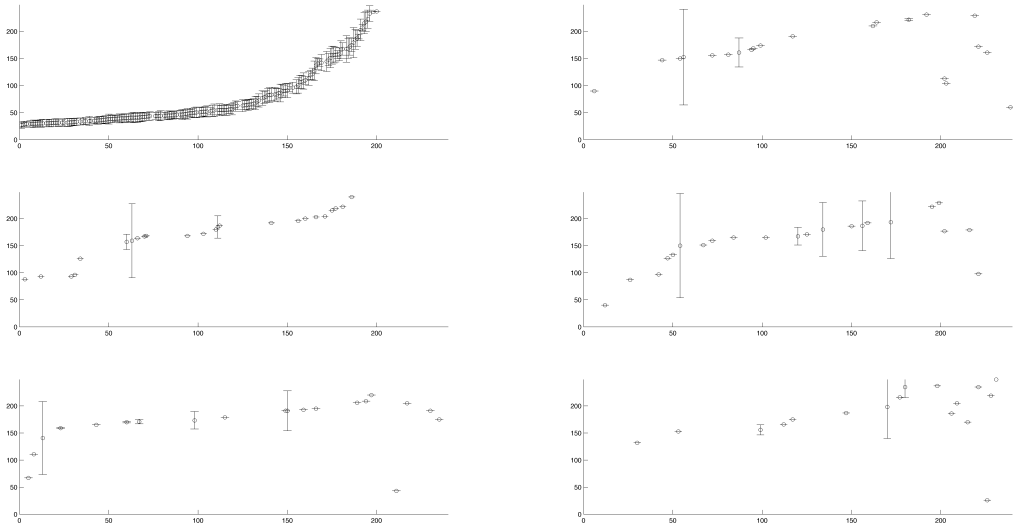

Figure 51: This plot shows the mean interspike interval for each neuron and its standard deviation, and has just 341 spikes. This is the lowest standard deviation run, and hence is found as the blue dot in bottom-left corner of Figure 50 (Internal Delay Range 4, External Delay Minimum 1, External Delay Range 16, Internal Connection Density 14, External Connection Density 8, with a Stimulation Frequency of 35 Hz).

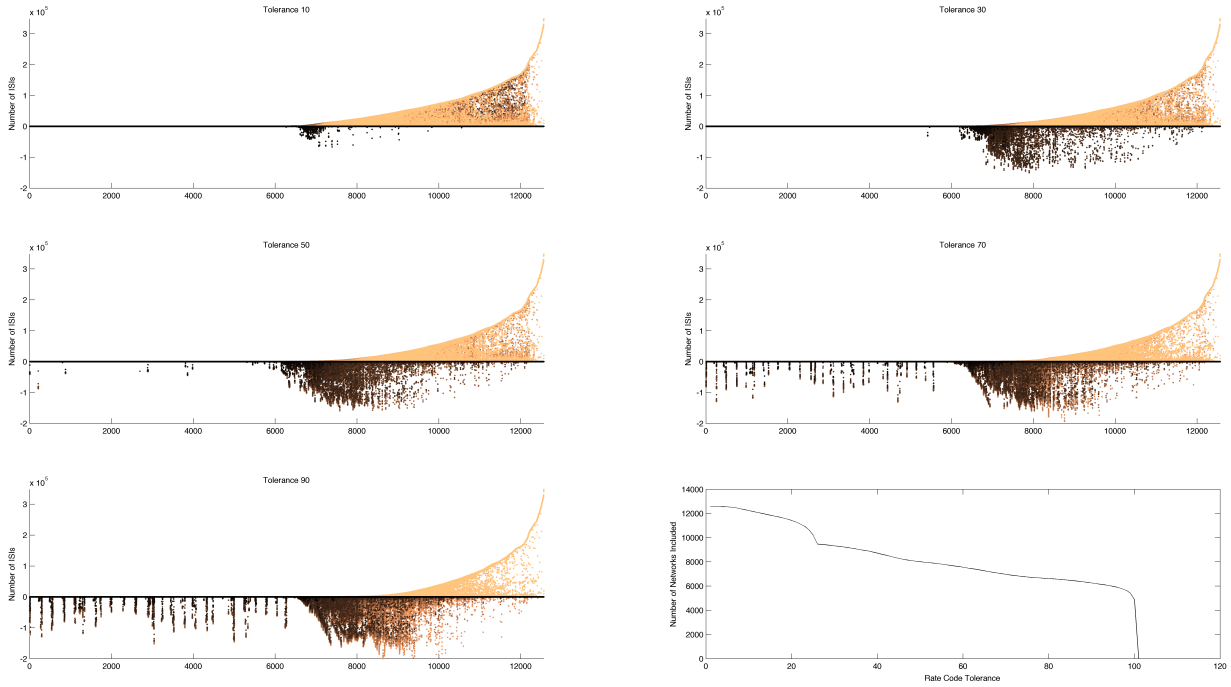

Figure 52: These plots show the way changing the Rate Code Tolerance impacts the number of interspike intervals to be included in the analysis, shown vertically, where each plot corresponds to a particular Rate Code Tolerance. Each plot is horizontally plotted relative to the number of interspike intervals to be included for each network, corresponding to rank of inclusion for a particular network. In a single plot, each network is vertically plotted as a dot colored twice per Rate Code Tolerance value from 1–100, where one dot is positively placed according to the number of interspike intervals to be included, and one dot is negatively placed according to the number of interspike intervals to excluded (overlapping dots have prioritized color according to the least inclusive Rate Code Tolerance for the range). The bottom right plot counts the number of networks with interspike intervals included, plotted across different values of RCT.

6.4 Parametric Analysis of Rate Code Stability

Using only data with Histogram Spectra Metric < 2000 that has over 200 spikes, interspike intervals for a neuron were extracted for analysis if the number of interspike intervals was greater than $\frac{30}{100}$ of its mean interspike interval, scaled by the size of the time window of observation. Statistics for each parameter value were then calculated from each included set of interspike intervals generated from networks employing that parameter value. On the left within Figure 53, the average of the mean of each interspike interval associated with horizontally plotted values of Stimulation Frequency, presented with its standard error. On the right within Figure 53 is the mean standard deviation of interspike intervals in conditions associated with each value of Stimulation Frequency—also presented with its standard error. An identical method was used to create each plot from Figure 53–58. These plots are intended to be used to understand the stability of rate coded information in the networks they describe. A network with a low average interspike interval standard deviation can be seen as having a a more identifiable mean interspike interval, requiring fewer rhythmic spikes to convey the mean. The mean interspike interval then describes the magnitude of the rate coding. The frequency with which each parameter value appeared in the set of included network simulations is listed in Table 6. This frequency information is important for understanding the relationships implied by the data.

Increasing *Stimulation Frequency* has the intuitive impact of decreasing the mean of interspike intervals within the network. Interestingly, though, as this Stimulation Frequency increases, so too does the standard deviation of the interspike intervals increase. This is counter to expectation, as the Poisson spike train stimuli have standard deviations proportional to the mean of the interspike intervals. This effect is best understood by considering the frequency of parameters that make up each point. The right most point, when Stimulation Frequency is 45 Hz, has a parameter distribution similar to Table 4. In this point distribution, ECD is always 8, and timing parameters EDR, EDM, and IDR are all biased towards sparsity. The lower ECD, combined with the increased sparsity of spike receipt, results in networks that can see larger amounts of stimulation without exploding. This also means that causally efficacious patterns will not be reliant on synchrony, resulting in higher standard deviations of the ISI, despite of having a lower mean ISI. When Stimulation Frequency is 35 Hz, however, the distribution across these parameters is completely flat—around 124 runs of each EDR, EDM, IDR, and ICD. When Stimulation Frequency is 25 Hz, the distribution of these parameters is similar to that of when it was 45 Hz, though instead of 355 out of 355 runs of ECD at 8, the 25 Hz has 510 out of 510

runs of ECD at 11, while also supporting higher ICD.

A higher *External Connection Density* leads to increased coupling across clusters, leading to increased synchrony and lower standard deviations of ISIs. Much like in the SF data, Figure 54 exhibits the same relationship between mean ISI and mean standard deviation of ISIs, though in a reverse parametric direction for ECD! This relationship is forced by the parameter distributions, as described for SF. Since each level of SF distinctly implies an opposite ECD value, the patterns mirror each other. The complexity of the interaction is further reduced from the restrictions on ECD, with $\frac{1484}{1628} = 0.9115$ proportion of the runs being contained in the conditions where ECD is 8 or 11.

External Delay Range, seen in Figure 55, has the impact of reducing the synchrony of receipt of intercluster spikes, yielding higher ISI means with increase. This effect is muted by the propensity for low EDR to amplify explosions, resulting in fewer unpruned runs with more gradually shifting dynamics. Higher EDR sees the average standard deviations of the ISIs trend down, an effect related to higher EDR enabling higher ECD through decreased synchrony of spikes upon receipt; the ratios of ECD 8 to ECD 11 runs shifts as EDR gets larger:

$$\left[\frac{175}{80}, \frac{192}{99}, \frac{204}{110}, \frac{200}{107}, \frac{203}{114}\right] = [2.1875, 1.9394, 1.8545, 1.8692, 1.7807].$$

This enables the ECD standard deviation effect visible in Figure 54 to proportionally shift the mean standard deviations of the ISIs for EDR.

Interspike interval means for *External Delay Minimum*, in Figure 56, appear uncorrelated with changes in EDM, as expected from how reliably it produces identifiable patterns of hypersynchrony. The interaction between EDM and low EDR or high ICD provides a pivoting point between its explosive parametric frequencies, balancing most parametric standard deviation interactions while being indifferent to ISI mean (its strongest ISI effects are identified as explosive for this analysis).

Increasing *Internal Connection Density* has the counterintuitive impact of increasing mean ISI in Figure 57, as a direct result of its reliability of causing explosive firing patterns. The difference between ICD of 5 (390 runs) and ICD of 17 (245 runs) is linear across all parameters—increasing ICD eliminates 16 different ECD 17 runs and 99 SF 45 runs—providing a clear source for the increase in mean ISIs observed here, as well as the unusual interactions seen in its standard deviation means.

Internal Delay Range is likely the least interactive parameter with respect to ISI statistics. Its ranges, from the expected zero variability cluster dynamics, to the dissipative 16ms delay ranges, play

	1	2	3	4	5
ICD	390	381	331	281	245
IDM	1628				
IDR	316	318	326	336	332
ECD	0	974	510	75	69
EDM	278	327	334	342	347
EDR	277	313	346	335	357
SF λ	0	144	510	619	355

Table 6: Distribution of parameter settings that construct an active network included in the interspike interval analysis.

cleanly with other parameters, operating on those parameters while allowing for lightly higher levels of activity without explosive drawbacks. As a result of this malleability, it sees no clear trend in this aggregation.

6.5 Rate Code Implications

While there are clear relationships between network structure and ISI stability, the overall magnitude of mean standard deviations of those ISIs imply the unreliability of mean ISIs. If these mean ISIs are taken to encode information, then that information is obscured by the extent of their variability. This result suggests that rate coded stimulation of clustered neural networks does not guarantee rate coded dynamics in downstream cells.

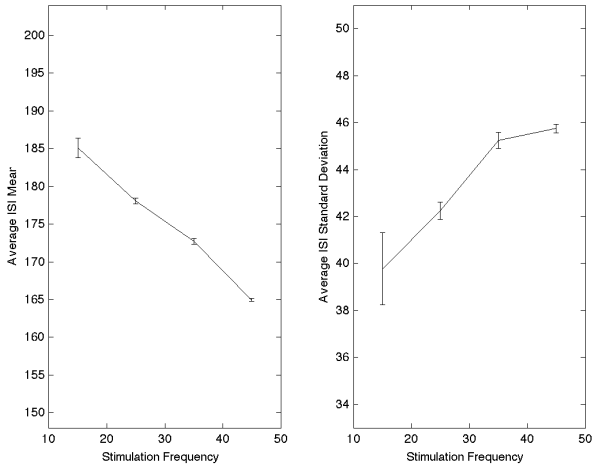

Figure 53: **Stimulation Frequency**

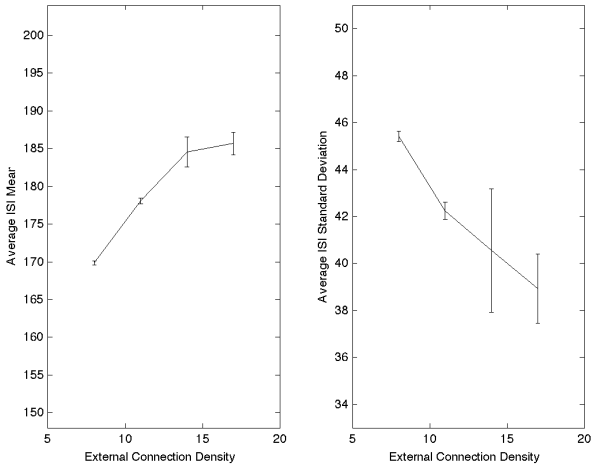

Figure 54: **External Connection Density**

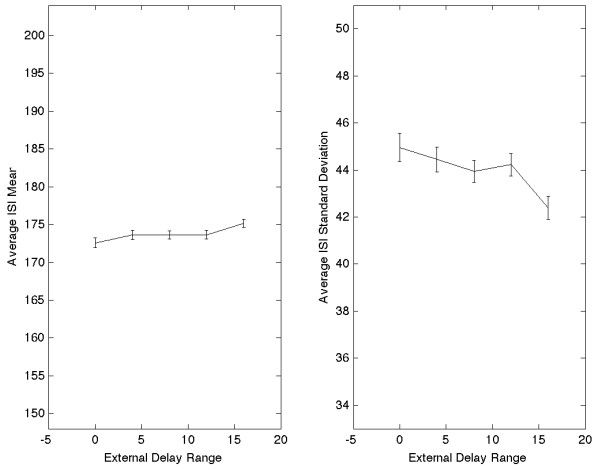

Figure 55: **External Delay Range**

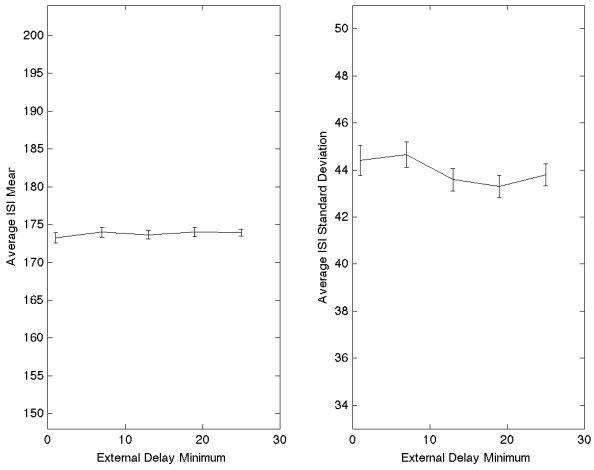

Figure 56: **External Delay Minimum**

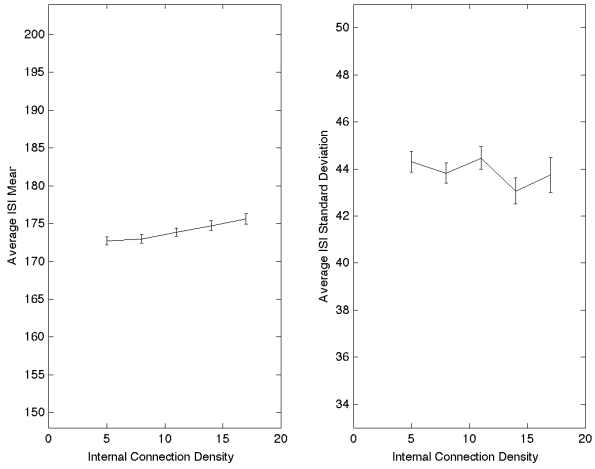

Figure 57: **Internal Connection Density**

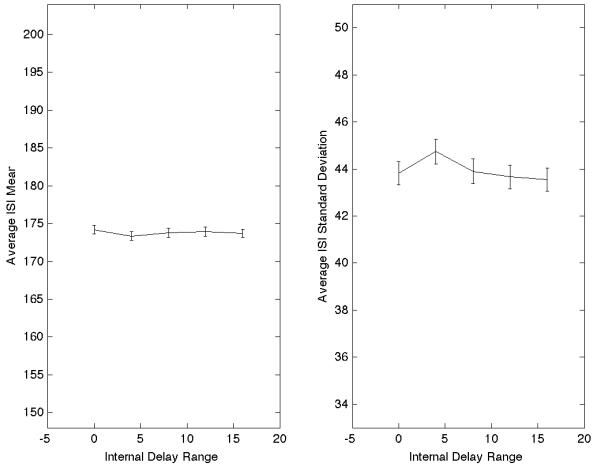

Figure 58: **Internal Delay Range**

7 Conclusion

The mystery of emergence reflects the frame of observation. The structure of that frame constitutes the historico-theoretical context within which phenomena assemble. To de-mystify the emergence of significant neural dynamics, we must consider those dynamics as supervening onto a space where significant dynamics appear imperative. Neural systems exhibit sensitivy to many scales. The complexity of these interactions is a testament to the importance of integrating constraints across scales to simplify the visualized dynamics. These constraints inform the validity of assumptions which surround the context wherein a model is seen as embedded. In construction, the impact of underlying assumptions become amplified as the scale and extent of interaction affects the shape of emergent phenomena. When these phenomena are considered as intertheorhetically supervening, however, their interpretation takes on the constraints of the overlying theoretical narrative. Sensitivity of an encompassing system across scales enhances the significance of these constraints, as nonlinearities exaggerate their impact.

In neural systems, the significance of local dynamics is offloaded into how it constrains the emergence of dynamics in the regions to which it projects. However, the small world connectivity structure of mammalian brains imply that local dynamics may come to influence what ultimately emerges over time. This relationship establishes the constraint between the long-term, multi-second behavior of the full system and the *meaning* of local dynamics, a notion which reveals the complexity of a genuine characterization of the neural code. While this sensitivity to conditions must be preserved to cope with

novel contexts, it is only *selectively* preserved. Similar initial dynamics can result in the emergence of similar downstream dynamics through bottlenecks among stable heteroclinic channels. Thus, the complexity of the neural code may be simplified by expanding the description of a neural dynamic through the heteroclinic channels through which it may weave.

In their simplest prescriptive expansion, a causally related set of spikes form a *polychronous neuronal group*, defined in terms of the triggers which enable its emergence from the network. The significance of PNGs can be extended both existentially, μ^E, and causally, μ^C, in a way that is compositionally meaningful, dependent on network topology, and context sensitive. The phenomenal flexibility of their characterization allow for the consideration of groups with arbitrary size and length, which, combined with their compositionality, provide a lens to observe the structure of significant dynamics across all scales. Critically, this observation provides a method to relate whole brain dynamical constraints to the smaller scale dynamics they existentially imply. Claims that apply in general to PNGs inform dynamics across many scales, providing utility as a faithful account of neural coding.

The fundamental spike composition of PNGs preserves their decomposibility back into those spikes, which likewise supports their compositional reorganization, allowing PNG dynamics to reinforce the trajectories that underly rate coded dynamics. This compositionality of PNGs provides a frame to evaluate supervening coding schemes, such as rate codes, particularly in terms of their impact on a regular spiking context. PNGs emerge, however, in neural networks with variability in delays of spike propagation. In past work, alternative coding schemes were cast as dichotonomous, where rate coding as opposite of synchronous coding or temporal coding schemes (Kumar, Rotter, & Aertsen, 2010). PNGs, however, employ a narrative that offloads the burden of coding scheme onto network architecture and manner of stimulation, yielding flexibility with any emerging dynamic. The effectiveness of rate coding, synchronous coding, or temporal coding, becomes a claim to predictiveness specific to the properties of the subsystems of interest.

Network structure determines critical points in the network, which changes what dynamics can be propagated without catastrophic interference (Section 5). In a context of a tonic rate coded stimulus, networks with larger ranges in moderately ranged delay enabled the stability of more densely connected networks by increasing their modularity through sparsity of receipt (Section 6). Within a densely connected, small world brain structure, these results suggest that the naturalistic asymmetry between distally projecting cortico-cortical patches with locally inhibited nucleic dynamics may provide utility

for the emergence of significant emergent dynamics. Furthermore, considering clustered networks with delays that are too short may facilitate catastrophic hypersensitivity, obscuring near-future information and establishing metabolistically dangerous network interactions. Importantly, these results suggest that the rate coded firing of an excitatory neuron in a densely connected cortical network cannot be assumed simply because it received rate coded synaptic stimulation.

These simulations, however, held synaptic strengths equal and constant, with the interest of limiting variability to stimulation and connectivity patterns. Patterned reinforcement of tuned synaptic strengths may be essential for the emergence of rate code stability. Neurons also did not have regulatory endocannabinoid systems, which are responsible for adapting the extent of neurotransmitter release to reduce post-synaptic excitability over longer time scales (Branco & Staras, 2009). While this study was limited in scope, it provides insight into important issues surrounding the emergence of neural firing patterns. Importantly, the simulations in Section 6 did not include bursting neurons, like those in prefrontal cortex or those which guide thalamic rhythms. These results are consistent with the interpretation that distal bursting projections may constrain the emergence of dynamics without excessively prescribing their response.

References

Abeles, M., Hayon, G., & Lehmann, D. (2004). Modeling compositionality by dynamic binding of synfire chains. *Journal of computational neuroscience*, *17*(2), 179–201.

Abney, D. H., Dale, R., Yoshimi, J., Kello, C. T., Tylén, K., & Fusaroli, R. (2014). Joint perceptual decision-making: a case study in explanatory pluralism. *Frontiers in Psychology*.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*, *9*(1), 357–381.

Alexander, S. (1920). *Space, time, and deity*. Macmillan.

Atmanspacher, H. (2011). Identifying mental states from neural states under mental constraints. *Interface Focus*.

Averbeck, B. B., Lehman, J., Jacobson, M., & Haber, S. N. (2014). Estimates of projection overlap and zones of convergence within frontal-striatal circuits. *The Journal of Neuroscience*, *34*(29), 9497–9505.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47–89.

Bak, P., Chen, K., & Creutz, M. (1989). Self-organized criticality in the game of life. *Nature*, *342*(14).

Barth, A. L., & Poulet, J. F. (2012). Experimental evidence for sparse firing in the neocortex. *Trends in neurosciences*, *35*(6), 345–355.

Bedau, M. (1997). Weak emergence. *Philosophical Perspectives on Mind, Causation, and World*, *11*, 375-399.

Beer, R. (2004). Autopoiesis and cognition in the game of life. *Artificial Life*, *10*(3), 309-326.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 640.

Bienenstock, E. (1995). A model of neocortex. *Network: Computation in neural systems*, *6*(2), 179–224.

Blackwell, R. H., & Schlosberg, H. (1943). Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *Journal of Experimental Psychology*, *33*(5), 407.

Block, N. (1997). Semantics, conceptual role. *The Routledge Encylopedia of Philosophy*.

Branco, T., & Staras, K. (2009). The probability of neurotransmitter release: variability and feedback control at single synapses. *Nature Reviews Neuroscience*, *10*(5), 373–383.

Broad, C. (1925). *The mind and its place in nature*. Routledge and Kegan Paul.

Butterfield, J. (2012, February). Laws, causation and dynamics at different levels. *Interface Focus*, *2*(1), 101-114.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge University Press.

Cessac, B., Paugam-Moisy, H., Viéville, T., & et al. (2010). Overview of facts and issues about neural coding by spikes. *Journal of physiology, Paris*, *104*(1), 5.

Cheour, M., & et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5).

Chorley, P., & Seth, A. K. (2011). Dopamine-signaled reward predictions generated by competitive excitation and inhibition in a spiking neural network model. *Frontiers in computational neuroscience*, *5*.

Churchland, P. (1989). *The neurocomputational perspective: Philosophical essays on the mind, brain, and science.* Cambridge, USA: MIT Press.

Churchland, P. (1996). *The engine of reason, the seat of the soul: A philosophical journey into the brain.* Cambridge, MA: MIT Press.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, *23*(3), 371–414.

Dale, R. (2008). The possibility of a pluralist cognitive science. *Experimental and Theoretical Artificial Intelligence*, *20*(3).

Dan, Y., Poo, M., & et al. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron*, *44*(1), 23–30.

Driesch, H. (1909). *Science and philosophy of the organism.* University of Aberdeen.

Driesch, H. (1914). *Problem of individuality.* Macmillan.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of neurophysiology*, *83*(3), 1733–1750.

Engel, A. K., König, P., Kreiter, A. K., Schillen, T. B., & Singer, W. (1992). Temporal coding in the visual cortex: new vistas on integration in the nervous system. *Trends in neurosciences*, *15*(6), 218–226.

Estany, A. (2001). The thesis of theory-laden observation in the light of cognitive psychology. *Philosophy of Science*.

Fant, G. (1960). *Acoustic theory of speech production.* Mouton and Co.

Fodor, J. (1974). Special sciences. *Synthese*, *28*, 97-115.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1), 3–71.

Fu, Q., Flament, D., Coltz, J., & Ebner, T. (1995). Temporal encoding of movement kinematics in the discharge of primate primary motor and premotor neurons. *Journal of Neurophysiology*, *73*(2), 836–854.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of neurophysiology*, *61*(2), 331–349.

Fuster, J. M., Alexander, G. E., et al. (1971). Neuron activity related to short-term memory. *Science*,

173(3997), 652–654.

Gärdenfors, P. (2000). *Conceptual spaces-the geometry of thought.* Cambridge, MA: Bradford Books, MIT Press.

Gardner, M. (1970). Mathematical games: The fantastic combinations of john conway's new solitaire game 'life'. *Scientific American*, *223*.

Gilbert, A. L., Regier, T., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *PNAS*, *103*(2), 489-494.

Goldman-Rakic, P. (1995). Cellular basis of working memory. *Neuron*, *14*(3), 477–485.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. *Comprehensive Physiology*.

Gonzalez-Burgos, G., Kröner, S., Krimer, L., Seamans, J., Urban, N., Henze, D., ... Barrionuevo, G. (2002). Dopamine modulation of neuronal function in the monkey prefrontal cortex. *Physiology & behavior*, *77*(4), 537–543.

Gordon, E. (2000). *Integrative neuroscience: Bringing together biological, psychological and clinical models of the human brain.* Harwood Academic Publishers.

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of experimental psychology*, *51*(1), 79.

Harnad, S. (1990). *Categorical perception: The groundwork of cognition.* Cambridge University Press.

Hayon, G., Abeles, M., & Lehmann, D. (2005). A model for representing the dynamics of a system of synfire chains. *Journal of computational neuroscience*, *18*(1), 41–53.

Henze, D. A., González-Burgos, G. R., Urban, N. N., Lewis, D. A., & Barrionuevo, G. (2000). Dopamine increases excitability of pyramidal neurons in primate prefrontal cortex. *Journal of neurophysiology*, *84*(6), 2799–2809.

Hodgkin, A., & Huxley, A. (1952a). Measurement of current-voltage relations in the membrane of the giant axon of loligo. *The Journal of physiology*, *116*(4), 424–448.

Hodgkin, A., & Huxley, A. (1952b). Propagation of electrical signals along giant nerve fibres. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 177–183.

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

Huxter, J., Burgess, N., & O'Keefe, J. (2003). Independent rate and temporal coding in hippocampal

pyramidal cells. *Nature*, *425*(6960), 828–832.

IPA. (1999). Phonetic description and the ipa chart. In *Handbook of the international phonetic association: a guide to the use of the international phonetic alphabet.* Cambridge University Press.

Izhikevich, E. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*.

Izhikevich, E. (2006). Polychronization: Computation with spikes. *Neural computation*, *18*(2), 245–282.

Izhikevich, E. (2007). Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cerebral Cortex*, *17*(10), 2443–2452.

Izhikevich, E. M. (2000). Neural excitability, spiking and bursting. *International Journal of Bifurcation and Chaos*, *10*(06), 1171–1266.

Izhikevich, E. M. (2007). *Dynamical systems in neuroscience.* MIT press.

James, W. (1890). The principles ofpsychology (vol. 1). *New York: Holt*.

Jin, D. Z., Fujii, N., & Graybiel, A. M. (2009). Neural representation of time in cortico-basal ganglia circuits. *Proceedings of the National Academy of Sciences*, *106*(45), 19156–19161.

Kayser, C., Logothetis, N. K., & Panzeri, S. (2010). Millisecond encoding precision of auditory cortex neurons. *Proceedings of the National Academy of Sciences*, *107*(39), 16976–16981.

Kellert, S., Longino, H., & Waters, C. (Eds.). (2006). *Scientific pluralism, studies in the philosophy of science.* University of Minnesota Press.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, *95*, 3-36.

Kriete, T., Noelle, D. C., Cohen, J. D., & OReilly, R. C. (2013). Indirection and symbol-like processing in the prefrontal cortex and basal ganglia. *Proceedings of the National Academy of Sciences*, *110*(41), 16390–16395.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27.

Kuhl, P. K. (2004, November). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831-843.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* University of Chicago Press.

Kumar, A., Rotter, S., & Aertsen, A. (2010). Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nature reviews neuroscience*, *11*(9), 615–

627.

Leibniz, G. (1714 and 1898). *Leibniz: The monadology and other philosophical writings* (R. Latta, Trans.). Oxford.

Levitt, J. B., Lewis, D. A., Yoshioka, T., & Lund, J. S. (1993). Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). *Journal of Comparative Neurology*, *338*(3), 360–376.

Levy, W., & Baxter, R. (1996). Energy efficient neural codes. *Neural Computation*, *8*(3), 531–543.

Lisman, J. E., Fellous, J.-M., & Wang, X.-J. (1998). A role for nmda-receptor channels in working memory. *Nature neuroscience*, *1*(4), 273–275.

Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nature neuroscience*, *4*(11), 1131–1138.

Macdonald, D., Creel, S., & Mills, M. (2004). *Biology and conservation of wild carnivores*. Oxford University Press.

Madhavan, R., Chao, Z., & Potter, S. (2007). Plasticity of recurring spatiotemporal activity patterns in cortical networks. *Physical biology*, *4*(3), 181.

Malisoff, W. M. (1939). Emergence without mystery. *Philosophy of Science*, *6*(1).

Martinez, R., & Paugam-Moisy, H. (2009). Algorithms for structural and dynamical polychronous groups detection. *Artificial Neural Networks–ICANN 2009*, 75–84.

McClelland, J., & et al. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*, 348-356.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing speech. *Nature*, *264*, 746-748.

Mill, J. (1843). *Systems of logic* (8th ed.). Longmans, Green, Reader, and Dyer.

Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & cognition*, *37*(7), 1026–1039.

Muro, C., Escobedo, R., Spector, L., & Coppinger, R. (2011). Wolf-pack (canis lupus) hunting strategies emerge from simple rules in computational simulations. *Behavioral Processes*, *88*, 192-197.

Nagel, E. (1961). *The structure of science*. Harcourt, Brace, and World.

Norris, J. R. (1998). *Markov chains* (No. 2008). Cambridge University Press.

O'Connor, T., & Wong, H. Y. (2012). Emergent properties. *The Stanford Encyclopedia of Philosophy*. Retrieved from `http://plato.stanford.edu/archives/spr2012/entries/properties`

-emergent

Oppenheiml, P., & Putnam, H. (1958). Unity of science as a working hypothesis. *The philosophy of science*, 405–427.

O'Reilly, R., & Frank, M. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural computation*, *18*(2), 283–328.

O'Reilly, R. C., Noelle, D. C., Braver, T. S., & Cohen, J. D. (2002). Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. *Cerebral Cortex*, *12*(3), 246–257.

Pasquale, V., Massobrio, P., Bologna, L., Chiappalone, M., Martinoia, S., & et al. (2008). Self-organization and neuronal avalanches in networks of dissociated cortical neurons. *Neuroscience*, *153*(4), 1354–1369.

Pavlov, I. P. (1927). *Conditioned reflexes*. DoverPublications.com.

Peterson, R., & Ciucci, P. (2003). The wolf as a carnivore. In *Wolves: Behavior, ecology and conservation* (p. 103-130). University of Chicago Press.

Poincaré, H. (1885). Sur l'équilibre d'une masse fluide animée d'un mouvement de rotation. *Acta mathematica*, *7*(1), 259–380.

Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science*.

Pucak, M. L., Levitt, J. B., Lund, J. S., & Lewis, D. A. (1996). Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. *Journal of Comparative Neurology*, *376*(4), 614–630.

Rabinovich, M., Volkovskii, A., Lecanda, P., Huerta, R., Abarbanel, H., & Laurent, G. (2001). Dynamical encoding by networks of competing neuron groups: winnerless competition. *Physical review letters*, *87*(6), 068102.

Rabinovich, M. I., Afraimovich, V. S., Bick, C., & Varona, P. (2012). Information flow dynamics in the brain. *Physics of life reviews*, *9*(1), 51–73.

Reinagel, P., & Reid, R. C. (2000). Temporal coding of visual information in the thalamus. *The Journal of Neuroscience*, *20*(14), 5392–5400.

Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, *73*(2), 713–726.

Rolston, J., Wagenaar, D., & Potter, S. (2007). Precisely timed spatiotemporal patterns of neural

activity in dissociated cortical cultures. *Neuroscience*, *148*(1), 294.

Ronald, E., Sipper, M., & Capcarrere, M. (1999). Design, observation, surprise! a test of emergence. *Artificial Life*, *5*, 225-239.

Rougier, N. P., Noelle, D. C., Braver, T. S., Cohen, J. D., & O'Reilly, R. C. (2005). Prefrontal cortex and flexible cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7338–7343.

Schrader, S., Diesmann, M., & Morrison, A. (2010). A compositionality machine realized by a hierarchic architecture of synfire chains. *Frontiers in computational neuroscience*, *4*.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*(4), 325–345.

Shepard, R. N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, *55*(6), 509.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. ii. *Psychometrika*, *27*(3), 219–246.

Shepard, R. N. (1965). Approximation to uniform gradients of generalization by monotone transformations of scale. *Stimulus generalization*, 94–110.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*(4468), 390–398.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological review*, *89*(4), 305.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Shepard, R. N., & Farrell, J. E. (1985). Representation of the orientations of shapes. *Acta Psychologica*, *59*(1), 103–121.

Spivey, M. (2008). *The continuity of mind* (Vol. 40). USA: Oxford University Press.

Sporns, O., & Zwi, J. (2004). The small world of the cerebral cortex. *Neuroinformatics*, *2*(2), 145–162.

Standish, R. (2001). On complexity and emergence. *Complexity International*, *9*.

St. Clair, W., & Noelle, D. (2013). Implications of polychronous neuronal groups for the nature

of mental representations. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 1372-1377.

St. Clair, W. B., & Noelle, D. C. (2015). Implications of polychronous neuronal groups for the continuity of mind. *Cognitive processing*, 1–5.

Steriade, M. (2004). Neocortical cell classes are flexible entities. *Nature reviews neuroscience*, *5*(2), 121–134.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Stuss, D., Floden, D., Alexander, M., Levine, B., & Katz, D. (2001). Stroop performance in focal lesion patients: dissociation of processes and frontal lobe lesion location. *Neuropsychologia*, *39*(8), 771–786.

Swadlow, H. (1985). Physiological properties of individual cerebral axons studied in vivo for as long as one year. *Journal of Neurophysiology*, *54*, 1346-1362.

Swadlow, H. (1988). Efferent neurons and suspected interneurons in binocular visual cortex of the awake rabbit: Receptive ?elds and binocular properties. *Journal of Neurophysiology*, *88*, 1162-1187.

Swadlow, H. (1992). Monitoring the excitability of neocortical efferent neurons to direct activation by extracellular current pulses. *Journal of Neurophysiology*, *68*, 605-619.

Szatmáry, B., & Izhikevich, E. (2011). Spike-timing theory of working. *BMC Neuroscience*, *12*(Suppl 1), O3.

Tabor, M. (1989). *Chaos and integrability in nonlinear dynamics: An introduction*. Wiley.

Trengove, C., van Leeuwen, C., & Diesmann, M. (2013). High-capacity embedding of synfire chains in a cortical network model. *Journal of computational neuroscience*, *34*(2), 185–209.

Vertes, P., & Duke, T. (2010). Effect of network topology on neuronal encoding based on spatiotemporal patterns of spikes. *HFSP journal*, *4*(3-4), 153–163.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*(6840), 953–956.

Whorf, B. L. (1956a). Science and language. In J. B. Carroll (Ed.), *Language, thought, and reality.* MIT Press.

Whorf, B. L. (1956b). Thought and reality. In *Language, culture and personality.* MIT Press.

Williams, G. V., & Goldman-Rakic, P. S. (1995). Modulation of memory fields by dopamine dl receptors in prefrontal cortex. *Nature*.

Wimsatt, W. C. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science*, *64*, S372-S384.

Wimsatt, W. C. (2006). Reductionism and its heuristics: Making methodological reductionism honest. *Synthese*, *151*(3), 445–475.

Xerol. (2009). *Glider trail.* Retrieved from `https://commons.wikimedia.org/wiki/File:Glider_trail.png` (License: `https://creativecommons.org/licenses/by-sa/3.0/deed.en` Attribution-Share Alike 3.0 Unported)

Yoshimi, J. (2012a). Active internalism and open dynamical systems. *Philosophical Psychology*, *25*(1), 1–24.

Yoshimi, J. (2012b). Supervenience, dynamical systems theory, and non-reductive physicalism. *British Journal of the Philosophy of Science*, *0*, 1-26.

Zipser, D., Kehoe, B., Littlewort, G., & Fuster, J. (1993). A spiking network model of short-term active memory. *The Journal of neuroscience*, *13*(8), 3406–3420.