A Complex Systems Approach to Understanding Cells as Systems and Agents

Moriah Echlin

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Ilya Shmulevich, Chair

Sui Huang

John Mittler

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

A Complex Systems Approach to Understanding Cells as Systems and Agents

Moriah Echlin

Chair of the Supervisory Committee:
Ilya Shmulevich
Department of Electrical and Computer Engineering

Many natural systems can be categorized as complex systems, with relatively simple components interacting to generate collective behaviors not easily predicted from the individual components themselves, like the flocking of birds or the formation of oceanic currents. Living systems, in particular, are enriched with complexity. In studying complex systems, abstract mathematical models are often used to identify general principles underlying how the interactions between individuals gives rise to observed collective behaviors. This type of approach allows for a focused investigation into the effects of specific lower-level properties (e.g., interaction distance) on higher-level behaviors (e.g., collective motion) in a controlled setting. In this work, I utilize Boolean network (BN) models to investigate cells, the fundamental units of life, as both systems of intracellular components and the agents that interact within cellular populations. Specifically, I

simulate cell-like agents composed of networks with binary-valued nodes. Agents can interact with their environment or with each other via external signals in the form of inputs to designated receptor nodes. With this model, I examine two overarching questions: (1) how internal variables influence the flexibility of cells to process external signals to generate different responses; and (2) how cell-cell communication impacts individual and population behavior in cellular populations. Using a BN reservoir computer model of cellular signal processing, I find that flexibility in signal processing is guaranteed if enough cellular resources (e.g., number of nodes) are available; however, fewer resources could attain flexibility, but with lower probability. I also find that the difficulty of accurately responding to signals is heavily dependent on how sensitive the response needs to be to signal variability. Using a 3D lattice-structured population of interdependent BNs as a model of cellular populations, I find that communication alone can induce cells to exhibit completely different sets of behaviors as compared with non-communicating cells. Furthermore, by tuning the distance over which cells can interact (interaction distance) and the amount of signal that activates a receptor (activation threshold), cellular populations exhibit distinct social behaviors, characterized by different cell type distributions and population diversity. Significantly, the maximum effects of cell-cell communication are observed when the interaction distance only includes one or two neighboring cells. Overall, in this work I have identified how key cellular properties relate to biologically relevant phenotypes, namely signal processing and self-organization.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my graduate advisor, Ilya Shmulevich. Thank you for your commitment as a mentor, for leading by example, and for being such a great person to know and work with. You introduced me to a field I did not know I was so passionate about and taught me how to contribute to it. If nothing else, I learned a bunch of Russian jokes.

Secondly, I would like to thank the rest of my graduate committee: John Mittler, Sui Huang, Carl Bergstrom, and Elhanan Borenstein. Thank you for the insight and feedback that helped guide me through the twists and turns of research.

Thirdly, I would like to thank the people who have supported me over the years of graduate school. To my fellow Ilya labbies and ISBers, thank you for listening, advising, encouraging, and providing food and caffeine. Thank you to Boris Aguilar and Dave Gibbs, in particular, who were my collaborators and helped me carry out the work contained here. To my friends and family, thank you for being wonderful people and enriching my life. To my husband, Jeff, thank you for always having my back.

Lastly, I would like to acknowledge the Molecular and Cellular Biology Program at the University of Washington and the Institute of Systems Biology in Seattle for institutional and financial support during my graduate work. I would also like to acknowledge the National Science Foundation; the material in this thesis is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1762114. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Chapter 1. INTRODUCTION

1.1 INTRODUCTION TO COMPLEX SYSTEMS RESEARCH

The world is full of highly involved, multidimensional phenomena, such as the movement of oceanic currents, the operations of electronic circuits, and the evolution of species. It is tempting to assume that the processes that lead to these phenomena are similarly intricate. However, humble origins can lead to grand outcomes. Interactions between simple units can give rise to complicated population behavior. For example, the simple eat/get eaten relationship between predator and prey can result in regular cycles in population dynamics [1,2]. Collectively, systems that exhibit such behavior are referred to as complex systems. Though there is no formal definition of complex systems, they can be loosely defined as systems of relatively simple agents that interact, giving rise to collective properties (or emergent behaviors) that are not readily predictable from the behavior of a single agent [3]. Given the minimal conditions for qualification, it is perhaps unsurprising that complex systems and their emergent behaviors are commonplace. From the synchronization of atomic spins in ferromagnets [4] to the generation of language in humans [4,5], complex systems encompass big and small entities, both living and nonliving. They are often nested, building upon each other - a system at one level of organization acting as a single agent in the system at the next level of organization. Remarkably, disparate systems display similar behavior despite a wealth of differences between them. One famous example of this is the appearance of Turing patterns in chemical reaction dynamics and developing embryos [6,7]. It is therefore thought that there are general underlying principles that govern the behavior of complex systems, such as the interaction between activating and inhibitory processes in the case of Turing patterns.

Because complex systems are so pervasive, there is strong research interest in understanding how the interactions between agents relate to emergent behaviors and what general principles might be behind those relationships. Much of the foundational work in complex systems research was led by physical scientists like Henri Poincaré (chaos theory), Alan Turing (pattern formation, computing), John von Neumann (computing), Murray Gell-Mann (quantum mechanics), Benoit Mandelbrot (fractals), Norbert Wiener (cybernetics), John Nash (game theory), and John Holland (genetic algorithms). Over time, however, complex systems research has expanded into the life sciences- e.g. biology, sociology, and ecology - as advances in computing power have made analysis of high dimensional, heterogeneous systems more tractable.

In order to study complex systems, researchers across both the physical and life sciences use tools borrowed from many disciplines. Some of the more heavily used tools come from the fields of dynamical systems theory, network theory, the theory of computation, and information theory [8–11]. The bias towards tools from these fields is likely in part due to the history of complex systems research in the physical sciences and more mathematically oriented disciplines like economics and evolutionary biology [12]. Regardless of the historical context, the generality of the concepts in these disciplines aligns well with the complex systems framework. It is possible to conceptualize most, if not all, systems as networks; to view interactions as the flow of information; and to consider dynamic processes as computation. Therefore, these tools help to bridge research across different domains and understand specific domains in a unified context.

1.2 CELLS AS COMPLEX SYSTEMS

The focus of this thesis is understanding cellular behavior using a complex systems framework. As the fundamental unit of life, cells are the simplest biological system, capable of exhibiting the

same basic behaviors as higher organisms. They obtain nutrients, respond to stimuli, and reproduce. A single cell can be an entire organism or a single piece of a larger organism. As such, insights from the study of cells are applicable across other levels of biology. Furthermore, because they are the fundamental unit of life, the inner workings of cells and how cells interact with each other underpins many fundamental questions in biology, such as how life evolved, how life transitioned from single to multicellular organisms, and how diseases manifest.

The majority of research investigating how cells function is reductionist in nature. This approach is very valuable and highly successful in learning how parts of cells work and how specific parts are related (directly or indirectly) to cellular or organismal behavior. However, living things are essentially hierarchies of nested complex systems [13]. Atoms interact to form molecules which have distinctive dynamics, such as the catalysis of a reaction by an enzyme. Macromolecules like proteins, fats, and nucleic acids form the building blocks of cells, interacting to create a cohesive living unit from nonliving parts. When cells interact they can become cellular aggregates or multicellular organisms that are capable of behaviors like division of labor or collective decision making. Organisms comprise populations that interact to form ecosystems or societies. At each level, new structures, properties, and behaviors appear that did not exist at the previous level. Thus, trying to understand the behavior of a higher level, e.g. cancer progression, by simply understanding the collection of behaviors exhibited at a lower level, e.g. genotype, is inherently an incomplete strategy [14]. For example, to follow the Turing pattern scenario, knowing that the protein wingless is involved in cell-cell communication during the differentiation of cells within a *Drosophila* embryo does not reveal the mechanism by which a striping pattern of cell types occurs or even that it should occur.

1.3 System-Based Methods for Understanding Cells

Recognizing the limitation in the reductionist method, there is great interest in studying cells as whole systems [14,15]. One of the most common systems approaches in biology is the statistical analysis of large experimental data sets. In this approach, large data sets are gathered from behaviorally different populations or different experimental conditions and analyzed for trends between the intracellular components and relevant phenotypes. For example, the relationship between gene expression and the stage of cancer progression can be considered with this approach. Note that it is still possible to take a reductionist approach with this methodology if intracellular components are treated independently. The systems approach, in contrast, investigates the behavior of several system components collectively, maintaining the relationships between them. The advantage of analyzing large experimental data sets in such a way is that the context of the entire system is considered when making connections between lower level properties and higher-level properties. However, because the scope under which data is collected is limited, extrapolation and interpolation to other system states is only possible with impractically large data sets [16]. In general, this approach relies on correlative analysis of data, primarily considering component interactions as a source of *a priori* knowledge. Because the role of component interactions in generating higher level properties is not considered, this type of research is not considered complex systems research, although it is undoubtedly a systems approach.

In order to explore the role of component interactions, an increasingly more prevalent approach is mechanistic modeling, which accounts for the specific interactions between system components at some level of abstraction through mathematical formalization [17–20]. Using formal analysis techniques or computer simulation, the relationship between the interactions of the components and their collective behavior is explored in a fully controlled setting. A key ideological

difference between statistical analysis and mechanistic modeling is that the former takes a given collective phenotype and identifies data trends in components that predict the phenotype while the latter takes given interactions between components and identifies the resultant collective phenotype. This does not mean that mechanistic modeling is devoid of data from real systems. Rather, there is a spectrum of mechanistic modeling approaches with varying degrees of data incorporation. On one end of the spectrum are heavily data-driven models that explore specific biological systems and on the other end are abstract mechanistic models that focus on a general concept without system-specific details.

As the name implies, data-driven mechanistic models are utilized for understanding mechanisms, interactions, and dynamic relationships in particular systems. In molecular and cellular biology, these types of models have been applied to the study of many cellular behaviors, including the cell cycle [21–23], immune system regulation [24,25], and quorum sensing in bacteria [26,27]. The advantage to this approach is that it can be used to generate predictions that enhance our understanding of that system's dynamics in a way that is applicable to and actionable in real systems. However, because the data used in generating these models is context- and system-specific, the models do not generalize well. For example, the findings from a model of the evolution of resistance in lung cancer are not likely to apply to breast cancer [28,29]. Additionally, increasing the dimensionality of the model makes both computationally or analytically exploring the model and interpreting the relationships between component interactions and phenotypes more difficult. Thus, while these models can make predictions about how a particular system may behave in a given circumstance, they often do not offer a greater understanding of complex systems as a whole.

On the other end of the mechanistic modeling spectrum, are abstract mechanistic models. In these models, many details are simplified or altogether ignored with the goal of preserving the core aspects of the system relevant to the question being posed. For example, modeling tumors as simplified populations of phenotypically-defined cells with differential fitness can provide insight into the effects of heterogeneity on the evolution of chemotherapeutic resistance [30]. This modeling approach allows for the exploration of general concepts that are applicable to a range of real systems at the cost of having limited predictive capacity for specific systems. Examples of other research in this area of modeling include the study of pattern formation [7,31], evolvability of populations [32–34], and collective computation [35,36]. Because many details are abstracted, the models remain tractable and relationships between lower-level properties with higher-level properties are more easily discernible. Moreover, at the current stage of complex systems research, there is still much to learn from "simple" models. Even starting with limited assumptions and basic agents, we are still unable to predict population behavior. From another point of view, these models can be utilized to learn what interactions or behaviors are sufficient to generate certain collective behavior. For these reasons, the work in this thesis is performed using abstract mechanistic models.

1.4 Boolean Networks as biological models

In this thesis, the dynamics of cells are described using the mathematical formalism of Boolean networks (BNs). Boolean networks are dynamic networks of binary-valued nodes interconnected by a set of Boolean update functions, each node with its own associated update function (Figure 1.1A). The network state changes over time through the application of the update functions, which take the value of a subset of other nodes at previous time steps as inputs (Figure 1.1B). Though simple, BNs are capable of rich dynamics, making them an attractive model for complex systems.

For the sake of clarity, I leave a detailed mathematical definition of Boolean networks and their dynamics to be included in the methods section of each research chapter.

Here, I will give a brief background of the use of BNs as biological models. Boolean networks were originally formulated in 1969 by Stuart Kauffman as models of gene regulatory networks [37]. Kauffman explored how simple randomly constructed networks could exhibit key properties of cellular dynamics, such as differentiation and homeostasis. He also used a similar model, the N-K model, to explore adaptive evolution over fitness landscapes [38,39]. Since 1969, Boolean networks have continued to be used as models of intracellular networks [40,41], including models of the yeast transcriptional networks [42,43], fly developmental modules [44], and mammalian stem cell differentiation [45,46]. In addition to gene regulatory networks, BNs have also been used as models of other types of systems, such as social [47-49], economic [50,51], robotic [52-54], and ecological systems [55,56]. Alongside BN models of specific systems, abstract BN models have been used to study general characteristics of intracellular and other complex networks, such as the system-level effects of certain classes of regulatory functions; the evolution and evolvability of networks; the size, structure, and scaling of dynamical steady states or cycles; the spread of perturbations; and the capacity for information storage [57-75].

One property in particular that is a recurring theme in Boolean network research, especially when applied to living systems, is that of criticality [76,77]. This property is not limited to BNs but applies to dynamical systems in general and is rooted in how perturbations spread within a system. On one end of the spectrum, there are ordered dynamics, in which perturbations to the network die out over time. On the other end there are chaotic dynamics, in which perturbations are amplified and spread throughout the network. Exactly at the transition between ordered and chaotic dynamics lies criticality [78]. Networks can be tuned to operate in these dynamical regimes via changes in key

parameters. For example, in the famous Ising model of ferromagnets, increasing the temperature of a ferromagnetic material will move the system from exhibiting ordered dynamics in which all atoms have the same up or down magnetic spin into chaos in which the atoms have random spins. Though they are not defining features of critical systems, criticality is associated with bifurcations in the dynamics of systems, maximized mutual information, and scale-free behavior involving power-law distributions [79]. In the case of the Ising model, this manifests as a switch between homogeneity and heterogeneity of spins (dynamic bifurcation), non-random spatial distribution of spins (maximized mutual information), and a distance-based power-law relationship between the spins of two atoms (scale-free behavior). Because of the unique properties of critical systems, it is thought that criticality is a common property of complex systems and a central property of living systems [80,81]. Therefore, many studies have focused on how criticality can arise in Boolean networks and what effects it has on network dynamics [82-86].

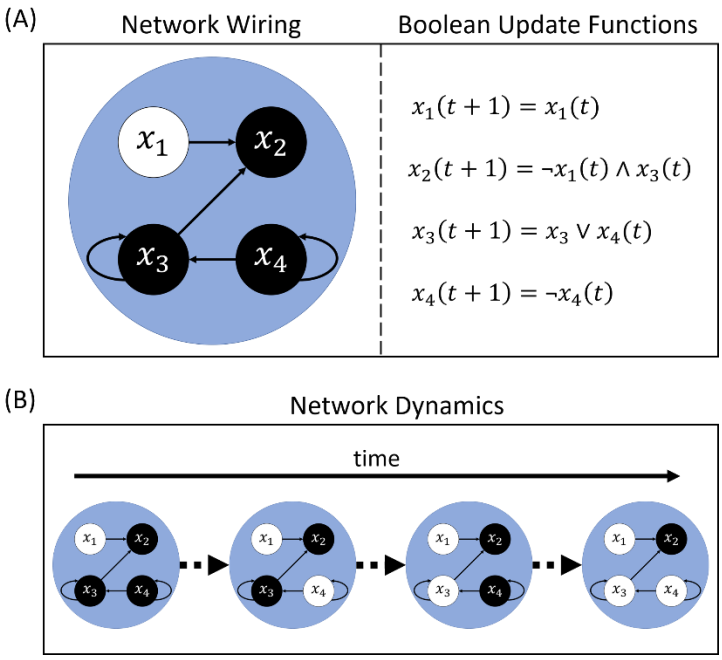


Figure 1.1. Diagram of cells as Boolean networks. **(A)** A cell (blue circle) is described by a network of variables (x_1, x_2, x_3 and x_4) that can take a value of either 1 (white) or 0 (black). The network is defined by a set of Boolean update functions (right panel), one for each variable, which take a subset of the variables as inputs. The functional dependence between network variables can be visualized as the wiring of edges between nodes (left panel). **(B)** By applying the update functions to the variables, the state of the network can be iterated through time. Boolean operations are denoted as such: AND, ∧; OR, ∨; NOT, ¬.

1.5 THE USE OF SIGNALS IN CELLS

In the following work, I use abstract Boolean network models to explore cells as both a system of interacting intracellular components as well as a single component in a greater system, i.e. cellular aggregates or multicellular organisms. I unify these two approaches by studying cells in the context of a greater environment, namely through the use of external signals. Whether the environment is abiotic or biotic, external signals are important for cells to tune in to their environment, using information from those signals to alter their behavior accordingly. In this way, cells are complex systems with many intracellular components interacting to modulate the response of the cell.

However, external signals can also serve as the interaction between cells as the units of a larger complex system - i.e. the cellular population - which will exhibit its own dynamics. Furthermore, in this work, I investigate the use of external signals by cells from two angles: (1) as a means of modulating behavior via internal machinery (Figure 1.2) and (2) as a form of interaction between complex cells within a population (Figure 1.3).

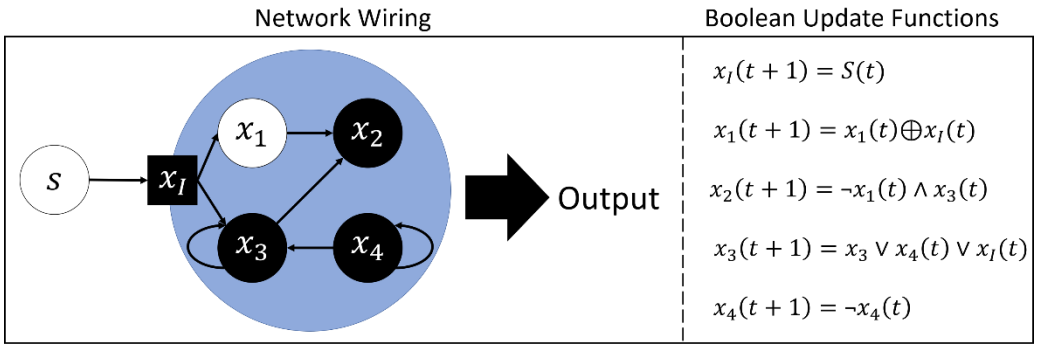


Figure 1.2. Diagram of BN cell responding to an external input. A cell (blue circle) senses the state of an external signal (S) via an input to a dedicated node (x_I) in the cellular network. The signal is processed by the cellular BN and the state of the network determines the response output by the cell. The Boolean update functions (right panel) and network wiring (left panel) reflect the cellular Boolean network's functional dependence on the external signal. Boolean operations are denoted as such: AND, $\wedge$; OR, $\vee$; NOT, $\neg$; XOR, $\oplus$.

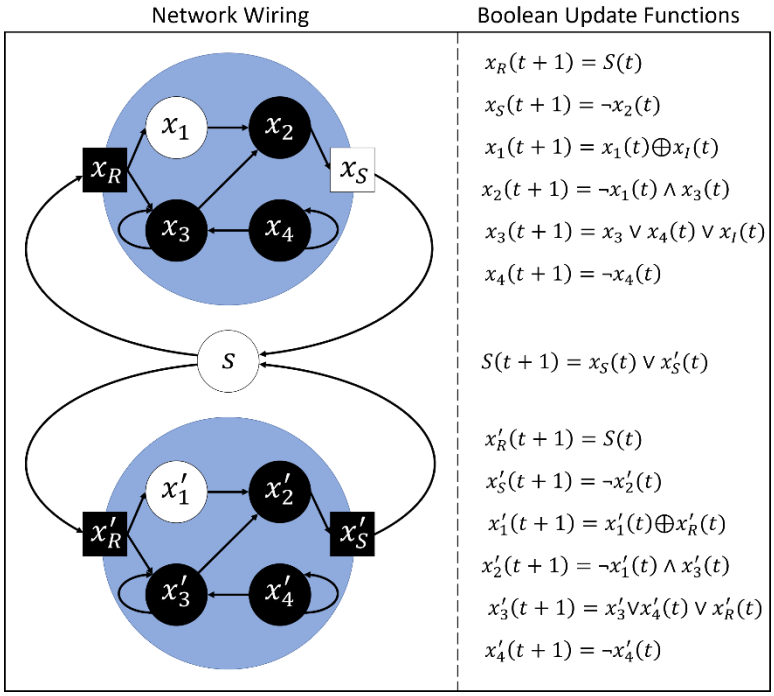


Figure 1.3. Diagram of two communicating BN cells. Each cell has its own corresponding Boolean network (x_1, x_2, x_3, x_4 and x'_1, x'_2, x'_3, x'_4). Communication between cells occurs via a common secreted signaling molecule node (S) which is secreted and sensed via wiring to dedicated cellular nodes (x_S, x'_S and x_R, x'_R, respectively). The Boolean update functions (right panel) and network wiring (left panel) reflect the cellular Boolean networks' functional dependence on each other through communication. Boolean operations are denoted as such: AND, ∧; OR, ∨; NOT, ¬; XOR, ⊕.

1.5.1 *Information Processing of External Signals in Cells*

The second chapter in this thesis is concerned with information processing in cells. As cells navigate the world, they are bombarded with signals that contain information about their surroundings. Based on these signals, cells must formulate a response. A straightforward example of this is the chemotaxis of *E. coli* toward sources of food. A bacterium senses the concentration of chemoattractant in its environment and uses the changes in that concentration to adjust its type of movement. By doing so, the *E. coli* bacterium moves in a directed random walk toward food.

Through the lens of complex systems, this can be considered an information processing task, in which the many components of a cell are interacting to perform the emergent behavior of processing an external signal to generate a behavioral change. Thus, utilizing the information contained in external signals is an important means by which cells modulate their behavior to adjust and adapt to environmental conditions, whether unicellular or multicellular.

Information processing, or the closely related concept of computation, by biological systems is not a new idea [9,87]. There has been research studying the capacity for information processing in real systems as well as research searching for principles and properties relevant to that capacity [88]. Advances in these areas are interconnected with our understanding of information processing; new tools and concepts in information theory and computer science can be applied to living systems. Over the last two decades, a new information processing framework has emerged that can be applied to (and inspired by) biology. That is the framework of reservoir computing [89,90] - a computational framework in which a signal is input into a reservoir (random nonlinear dynamical system) that processes the signal. The state of the reservoir is input into a function - typically simple, such as a linear one - to produce an output value. The output value is estimating some specified function of, or response to, the input. The key concepts here are that reservoirs are random nonlinear dynamical systems with fading memory that perform online signal processing tasks. This has a clear connection to intracellular networks, which can be modeled as semi-random nonlinear dynamical systems with fading memory that take in signals and "generate" output behaviors in continuous time. In fact, many other non-electronic physical systems have been shown to be capable of acting as reservoir computers, including systems of neurons [91]. Thus, in order to investigate the relationship between the intracellular components of a cell and its ability to process external signals, I studied an abstract cellular system as a reservoir computer [92]. This

builds on previous work in two areas: first, in the field of cellular information processing by considering temporal information processing with the reservoir computer framework; second, in the field of reservoir computers by considering biological reservoir computing in an abstract cell as well as computer-based reservoir computing with a Boolean network framework.

1.5.2 *Cells Interacting through Signals*

The third chapter of this thesis is concerned with the use of signals as a form of interaction between cells within a population. Signaling between cells is distinct from other forms of cell-cell interactions in that it is directed but not inherently linked to a specific cellular process. To clarify, signaling between cells is distinct from indirect interactions through the environment, such as competition for resources or modification of the environment. It is also distinct from direct interactions that either do not affect intracellular dynamics (e.g. mechanical forces) or affect intracellular pathways via vital cellular processes (e.g. metabolite exchange, antibiotic production). Rather, a signal sent by one cell is directly received by other cells and processed through pathways that have evolved in the context of the social environment. Not being tied to vital cellular processes, the signals can be adapted for other functional effects through evolution without major disruption to basic cellular operations. However, as with most other aspects of biology, the distinction can be fuzzy with overlap between these categories, such as intracellular signal transduction due to mechanical interactions or the adaptation of antibiotics to influence virulence factors.

There has been a significant amount of work in the complex systems field studying general properties of intracellular dynamics. However, the majority of this work has studied cells as isolated systems in terms of exchanging information with their environment [93-99]. Thus, not much is known about how time varying external input will change these dynamics. Similarly, a sizeable

amount of literature focuses on how signals between simple binary-valued cells (i.e., cellular automata) can affect population dynamics [31,100-103]. Therefore, I focus on bridging the gap between these two approaches with the aim of understanding the effect of signaling on the complex dynamics of cells as well as the effect of dynamically complex cells on cellular population dynamics, namely using information-theoretic metrics. This work builds on a handful of similar network-of-networks studies in Boolean networks, including the study of tissue patterning in sheets of cells [104], and behavioral diversity in tissues [105].

Chapter 2. FLEXIBILITY OF BOOLEAN NETWORK RESERVOIR COMPUTERS IN APPROXIMATING ARBITRARY RECURSIVE AND NON-RECURSIVE BINARY FILTERS

This chapter is adapted from Echlin et al. [92], previously published by *Entropy*, a publication of MDPI.

2.1 INTRODUCTION

Many biological systems are regarded as non-linear dynamical systems operating in high-dimensional space. Proteins, genes, and macromolecules interact in a variety of ways to create the dynamics of cells [89]. Collections of cells interact and coordinate activity, forming cohesive units such as bacterial colonies, simple multicellular organisms, or tissues in more complex multicellular organisms. At each level of organization, 'input' signals (e.g., odors or hormones) are introduced into the system, processed by means of the system's dynamics, and responded to accordingly, sometimes generating new 'output' signals as a byproduct [106]. Fundamentally, biological systems must process external signals in real time to inform a wide variety of response decisions. For example, the fruit fly olfactory system projects an input to a high dimensional space before classifying an odor [107].

In this vein, reservoir computing (RC) is a form of signal processing used for classification and online learning that uses such high-dimensional systems to process signals. Initially called echo state networks [89] or liquid state machines [90], the computational structure resembles an unorganized recurrent neural network. These networks were inspired by the most classical example of signal processing in biological systems — the brain. While initial research in neural networks was focused on modeling and understanding neural computation [108] applications and directions for

research quickly expanded in scope. For reservoir computers, this includes systems for predicting time series from chaotic systems [109,110] as well as multiple real-world applications such as gas detection [111], robotics [112-114], and economic trends [115]. Applications in processing human data are also found with human emotion [116] and gesture recognition [117], speech and text processing [118,119], and health care monitoring [120].

In biology, reservoir computers have been used to analyze biological data for identifying anomalous states such as cardiac arrhythmia [121], seizures [122,123], and microsleeps [124] as well as for high-dimensional classification such as cell type detection [125]. In these applications, RCs are used to interpret or classify biological data. RCs have also been used as models of biological systems, serving to explore the mechanisms and dynamics of said systems. For example, the questions of how the brain functions [126-128] and how gene regulatory networks respond to external stimuli [129,130] have been studied in an RC context.

In reservoir computing, an incoming signal is fed into the reservoir, which is a randomly connected recurrent network where nodes are connected via a variety of coupling functions. Thus, the reservoir transforms the signal into a high-dimensional representation. Finally, an output layer is trained, often with straightforward techniques such as regularized regression [131], in order to perform a classification or regression task (Figure 2.1).

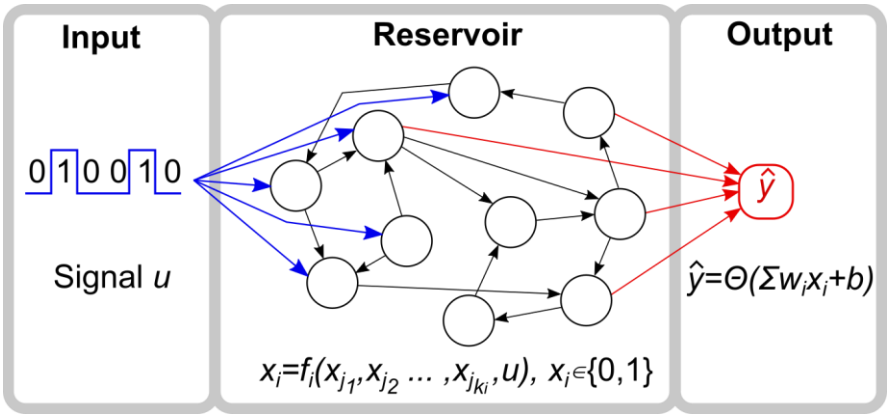


Figure 2.1. Reservoir computer layout. The RC is composed of a binary input node, a Boolean network reservoir, and a binary output node.

RCs can also be trained to act as signal processors, filtering the input to produce a new output through approximating a function that is normally applied to a window sliding over the signal. The type of function can be extended to recursive functions, where some elements of the input window are replaced by values from previously generated outputs. These types of filters have a long history of use which includes filtering biological signals [132-134].

While reservoirs are most typically built using unorganized recurrent neural networks, it has been shown that any non-linear dynamical system that exhibits a fading memory property can theoretically be used to compute a time-invariant function on a signal [135]. In fact, DNA, carbon nanotubes, memristor networks, and even buckets of water have been shown to work as reservoirs [136-138]. One type of dynamical system that has only recently been explored as a reservoir is Boolean networks [139]. Boolean networks (BNs) are well-suited to reservoir computing since they are one of the simplest modeling approaches that (1) can capture the heterogeneity, both in wiring and coupling functions, that characterizes self-assembled systems; and (2) can exhibit the non-trivial dynamical behavior that is required for computation [140]. Additionally, the fading memory property can easily be achieved in Boolean networks by adjusting two key parameters: the average in-degree

of the Boolean functions, $\bar{K}$, and the bias, p, which is the probability that a function produces an output of one for any given set of inputs [82,141]. Implementation of Boolean networks, both in software and in hardware, is also considerably easier and faster than more traditional approaches [142].

Similar to neural networks, Boolean networks were first used as models of biological systems, namely gene regulatory networks [139,143,144]. While often researched to understand cellular behavior in terms of protein/gene interaction networks in isolation [145-147], BNs have also been used to understand how external signals impact biological systems [105,148,149]. These systems rely on their ability to process external signals (relevant to survival or population function) and make decisions based on them in real time, thereby performing a similar computation to a reservoir [107]. Just as the Boolean network framework has provided insight into how biological systems function, studying BN reservoir computers (BN RCs) may further our understanding of biological signal-response.

Previous work on BNs as reservoirs is still limited. Snyder et al. [140,150] evaluated BN RCs with different numbers of nodes (N) and average node input degrees ($\bar{K}$). Using a measure of reservoir quality, it was found that, compared to networks with homogeneous in-degree, networks with heterogeneous in-degree better separated inputs by class: the so-called separability property [90]. Also, it was found that computational ability - when the difference in separability and fading memory is greatest - was maximized when the average in-degree was $\bar{K} = 2$ (out of $\bar{K}$=1,2,3) [150], and that higher $\bar{K}$ led to worse performance. This value for $\bar{K}$ is notable, as Boolean networks with $\bar{K} = 2$ are known to be dynamically critical (Lyapunov exponent = 0). Critical dynamics, which lie at the border between order and chaos, have been shown to be important in computation and information processing, biological and otherwise [80,85,151-153].

While previous work [140,150] has laid the foundation showing that it is possible to use Boolean networks as reservoirs, the main focus has been on testing a small number of well-known functions, such as the median and parity functions, as benchmark measures. We are extending the body of research in three ways: (1) by performing a wider examination of how well these systems perform across different functions, including recursively defined operators; (2) by analyzing how flexible a single reservoir is for repurposing (i.e., retraining only the output layer and keeping the same network reservoir); and (3) characterizing the functions that are easier or harder for the reservoir to implement in terms of the function's average sensitivity, which is a measure of its smoothness. We also evaluate reservoirs under different dynamical regimes: ordered, critical, chaotic.

2.2 METHODS

2.2.1 *Reservoir Computer*

A reservoir computer consists of three components: the input layer, the reservoir network, and the output layer (Figure 2.1). The input represents a temporally changing signal, u, that perturbs the reservoir, which performs computations on the signal in real time. The output node is the readout of the reservoir's computation, estimating the value, y, of a given function, $g(u)$, operating on the signal. The value of the output node, $\hat{y}$, is given by a linear combination of the states of the reservoir nodes that are continuously being driven, either directly or indirectly (via other nodes), by the input signal. The weights of the linear regression are trained to be specific to a given objective function, which captures the error between the reservoir output and the function to be approximated.

2.2.2 *Reservoir*

In our implementation, the reservoir is constructed as a random Boolean network (RBN) [139]. RBNs are networks with N binary-valued nodes with states

$$X_t = \{x_1{}^t, x_2{}^t, \ldots, x_N{}^t\} \qquad x^t{}_i \in \{0,1\}, i = 1, \ldots, N, t > 0 \tag{2.1}$$

The state of the network at any time, X_t, is a function of the state at the previous time, X_{t-1}, given by:

$$X_t = F(X_{t-1}) \tag{2.2}$$

where:

$$F = \{f_1, f_2, \ldots, f_N\} \tag{2.3}$$

is the set of Boolean updating functions corresponding to each node. The node in-degree does not need to be constant, such that each function f_i has an independent number of arguments, k_i, so that:

$$x_i{}^t = f_i(x_{j_1}{}^{t-1}, x_{j_2}{}^{t-1}, \ldots, x_{j_{k_i}}{}^{t-1}) \tag{2.4}$$

The identity of each of the k_i arguments is chosen randomly from the N nodes, without replacement. The in-degree of the whole network, $\bar{K}$, is characterized by the average in-degree across its nodes. The other primary descriptor of the network is the bias, p, which is the probability that a function outputs a one. Together, p and $\bar{K}$ can be varied to adjust the dynamics of the network, ranging from ordered, in which perturbations die out, to chaotic, in which perturbations are amplified [82,141]. For this paper, we will fix p= 0.5 and tune the dynamics of the networks by varying $\bar{K}$. Generally, we tune the dynamics using $\bar{K}$< 2 for ordered, $\bar{K}$= 2 for near critical, $\bar{K}$> 2 for chaotic. To construct a network with a specific $\bar{K}$, we take $\bar{K} \times N$ edges uniformly distributed amongst all pairs of nodes.

2.2.3 *Input*

In order for the RBN to act as a reservoir computer, it must be driven by an external signal, represented by a temporal sequence of binary values, u^t. Some percentage (expressed as a fraction) of reservoir nodes, L, are directly connected to the input signal. If a node, i, is connected to the input signal, then the input becomes an additional argument to its function, i.e.:

$$x_i{}^t = f_i(x_{j_1}{}^{t-1}, x_{j_2}{}^{t-1}, \ldots, x_{j_{k_i}}{}^{t-1}, u^{t-1}) \tag{2.5}$$

The $L \times N$ nodes that have the input as an additional argument are chosen uniformly from the N nodes of the reservoir without replacement.

2.2.4 *Output*

At each time step, the reservoir produces a binary output value, $\hat{y}^t$, defined as:

$$\hat{y}^t = \Theta(\sum_{j=1}^{N} w_j x_j{}^t + b), \tag{2.6}$$

where Θ is the function that maps values greater than 0.5 to 1, everything else to 0; w_j is a weight for each node of the reservoir and b is a constant. The parameters w_j and b are trained to approximate the output, y^t, of a given Boolean function, $g(u)$, operating on a moving window over u^t, under some error criterion. Here, we use every node in the reservoir in the linear regression so that an optimal combination may be found. Pre-specifying which nodes are used in calculating the output would limit the success of the reservoir, since the structure and rules of the reservoir are fixed. In practice, when N= 100 and L= 0.5, ~25% of the nodes have non-zero weights, which are distributed equally across the nodes that are directly and indirectly perturbed by the signal.

2.2.5 *Objective Functions*

In this work, the computational performance of the reservoir computer is assessed by approximating Boolean functions of three and five arguments evaluated over a temporally changing input signal. Moreover, we considered two types of functions:

- Non-recursive functions, defined as $y^t = g(u^{t-\tau}, u^{t-\tau-1}, \ldots, u^{t-\tau-(M-1)})$ (2.7)
- Recursive functions, defined as $y^t = g(u^{t-\tau}, u^{t-\tau-1}, \ldots, u^{t-\tau-(M-2)}, y^{t-1})$ (2.8)

where M is the length of the sliding window on the input signal for which a function is approximated, and τ is a delay between when the signal perturbs the reservoir and when the corresponding output is computed. In this work we explore M= 3,5 for non-recursive functions, and M= 3 for recursive functions. Here, τ is kept fixed to τ = 1. We have tested all 256 three-bit recursive and non-recursive functions, and 1000 randomly sampled five-bit functions.

2.2.6 *Training and Testing Algorithm*

To train a BN RC for approximating a single function, we use random binary sequences as input streams, u^t, and compare the reservoir's output value, $\hat{y}^t$, to the function output, y^t, evaluated over the input stream. Specifically, we use a set of 150 random binary sequences of length $10 + M - 1$ (e.g., 12 for 3-bit functions) to generate a set of P = 1500 different $(\hat{y}, y)$ pairs. It should be noted that for each $10 + M - 1$ binary sequence, the reservoir is randomly initialized and no transient period occurs before outputs are used for training. With the $P(\hat{y}, y)$ pairs, a regression model is fit in order to generate the coefficient weights on the output layer. We used Scikit-learn to perform lasso regression [154] with an α = 0.1 for fitting.

To test a BN RC for approximating a single function, g_j, we again use a set of 150 random binary sequences to generate $P(\hat{y}, y)$ pairs. We then measure the accuracy of the BN RC for that function, defined as:

$$a_{ij} = 1 - \rho / P \ , \tag{2.9}$$

where j indexes the Boolean function, i is a particular instantiation of an RC, and ρ is the summed error $\sum_{l=1}^{P} |\hat{y}^{\,l} - y^l|$, which is scaled by P. Thus, it can be seen that the accuracy a_{ij} is between 0-1. This process of training and testing is repeated for a set of different functions for each BN RC, thereby creating multiple estimation accuracies a_{ij}.

2.2.7 *Overall Strategy*

The goals of this work center around exploring how the approximation accuracy of RBN RCs varies across many different functions, providing a sense of 'flexibility'. To do that, one reservoir is constructed and applied to estimating the output of a set of Boolean functions, such as all three-bit functions. This is done for a total of 100 RC instances for each combination of different values of N, L, and $\bar{K}$ to test the effect of reservoir size, degree of network perturbation by the signal, and dynamical regime, respectively. We use N= 10, 20, …, 50, 100, 200, ..., 500; L= 0.1, 0.2, ... ,1; and $\bar{K}$= 1,2,3. For example, with the three-bit functions, we trained three (values of $\bar{K}$) × 10 (values of N) × 10 (values of L) × 100 (RCs instances) × 256 (three-bit functions) = 7,680,000 RBN RCs.

Each RC is trained and tested for approximating a set of functions, including median and parity, reporting an accuracy for each function. We chose three sets of functions for testing the RCs. First, there are the three-bit functions, for which we test all possible $2^{2^3} = 256$ functions. Second, there are five-bit functions, for which there are $2^{2^5} = 4{,}294{,}967{,}296$ functions, which is a much larger function space than for three bits. Working with the complete function set is

impractical, so we made a random sample over the function space, choosing to test 1000 functions. Some additional hand-selected key functions, such as median and parity, were also used.

The last set of functions was recursive three-bit Boolean functions. This set of functions is defined as taking a set of arguments that includes the last produced output, $y^t = g(u^{t-\tau}, u^{t-\tau-1}, y^{t-1})$. This is a more difficult task, because the BN RC must approximate the function with a hidden variable, y^{t-1}. Software for constructing and running BN RCs can be found at https://github.com/IlyaLab/BooleanNetworkReservoirComputers.

2.3 RESULTS

2.3.1 *Benchmark Functions: Median and Parity*

We start our analysis with the approximation of the two functions tested in previous BN RC works[140,150]: the temporal median and parity functions. The median function calculates whether there are more 1s than 0s in the window and is defined as:

$$d_t = \chi[\sum_{i=0}^{M-1} u(t-\tau-i) > M/2] \tag{2.10}$$

where $\chi[A]$ is an indicator function that gives a one if A is true and zero if A is false. The parity function calculates whether there is an odd number of ones in a bit string. It is defined as:

$$p_t = \bigoplus_{i=0}^{M-1} u(t-\tau-i) \tag{2.11}$$

where $\bigoplus$ is addition modulo 2. These two functions are frequently used as benchmarks for reservoir performance. For each function g_j, we trained 100 BN RCs for each set of the parameters $\bar{K}$, N, and L, measuring the mean accuracy with respect to the 100 BN RCs:

$$\bar{a}_j = \frac{1}{100}\sum_{i=1}^{100} a_{ij} \tag{2.12}$$

2.3.2 *Median*

On average, the three-bit median function is approximated with high accuracy (Figure S2.1). All BN RCs perform better than random, with the lowest mean accuracy of ~0.7 for N = 10 and L=0.1. Increasing N and L increases accuracy logarithmically up to ~0.98 for N= 500 and L= 1 (Figure 2.2A). However, reservoirs of size N > 200 already have near-perfect performance (>0.9), and increasing L has a minimal effect on these reservoirs. The general trend in the accuracy in relation to N and L remains the same for the five-bit median (Figure S2.2). However, the overall accuracy is lower relative to the three-bit median (Figure 2.3A).

The average in-degree, $\bar{K}$, of the reservoirs also affects the reservoir performance. For the three-bit case, the performance is clearly reduced for $\bar{K}$ = 3, with those reservoirs having the lowest accuracy. $\bar{K}$ = 1 and 2 have similarly high levels of accuracy (Figure 2.2A). However, increasing the size of the function window from three to five creates a more obvious separation between these $\bar{K}$ values (Figure 2.3A). Interestingly, for most values of N and L, the accuracy for the reservoirs where $\bar{K}$ = 2 perform the best; however for higher values of N and L, the accuracy for the reservoirs where $\bar{K}$= 1 are the highest. Overall, these results qualitatively agree with previous work by Snyder et al. [140,150].

Unexpectedly, when the reservoir is tasked with approximating a three-bit recursive median, the performance is nearly as high as for the non-recursive three-bit median, and higher than for the five-bit (non-recursive) median (Figure 2.4A). Two notable differences for the recursive median are that (1) increasing L does not grant as sharp of an increase in accuracy, and (2) the performance of $\bar{K}$ = 3 is more reduced as compared to lower $\bar{K}$ (Figure S2.3).

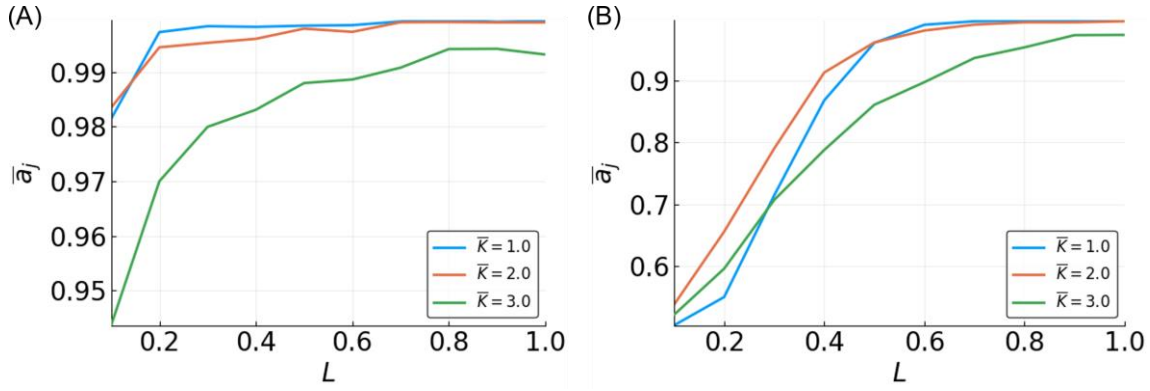


Figure 2.2. Mean accuracy, $\bar{a}_j$, vs. L for the three-bit median **(A)** and parity **(B)** functions for different $\bar{K}$-valued reservoirs with N = 500.

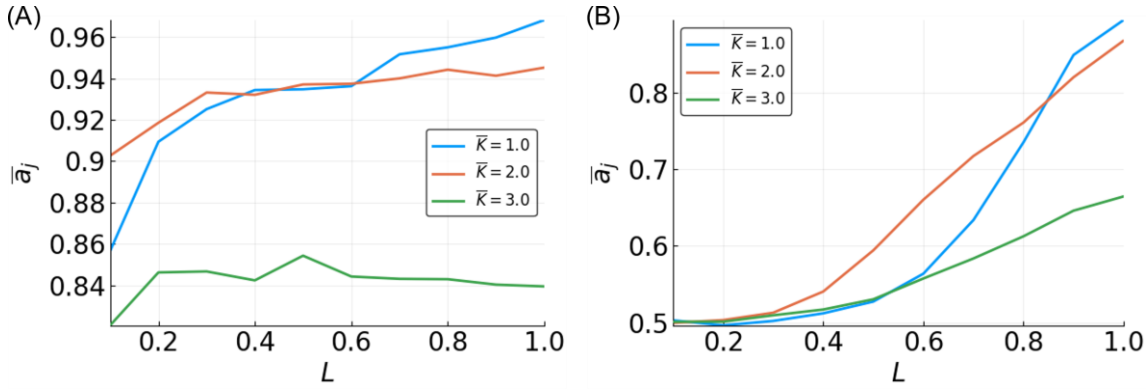


Figure 2.3. Mean accuracy, $\bar{a}_j$, vs. L for the five-bit median **(A)** and parity **(B)** functions for different $\bar{K}$-valued reservoirs with N = 500.

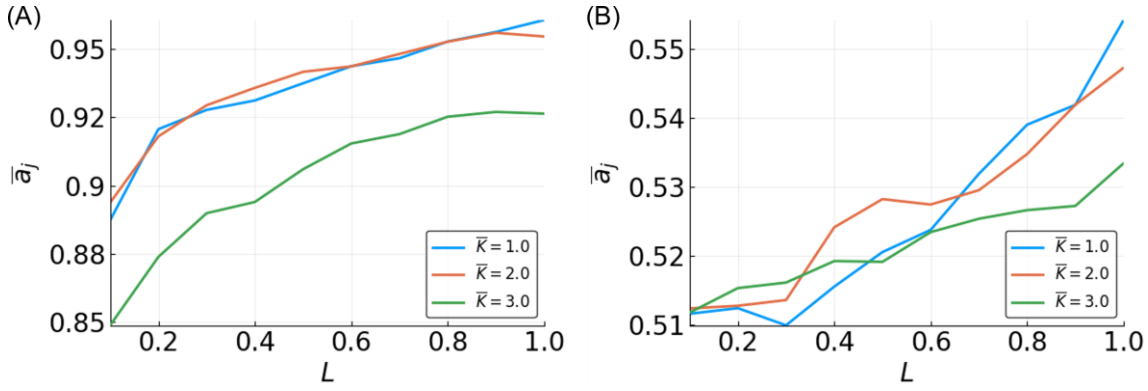


Figure 2.4. Mean accuracy, $\bar{a}_j$, vs. L for the recursive three-bit median **(A)** and parity **(B)** functions for different $\bar{K}$-valued reservoirs with N = 500.

2.3.3 *Parity*

The accuracy curves for approximating the three-bit parity function have a strikingly different appearance than those for the median (Figure 2.2B). Accuracy is highly dependent on the value of L, with no reservoir performing better than random (~0.5) for L = 0.1. Very small reservoirs (N<40) are incapable of high accuracy (Figure S2.1). As reservoirs become larger, the maximum accuracy increases. When the reservoirs are used to approximate the five-bit parity function, the performance drops significantly (Figure 2.3B). No reservoir is capable of performing better than random at L < 0.5, and an accuracy > 0.8 is only achievable with larger networks (N > 300) and L near 1 (Figure S2.2). These observations are consistent with previous work [140,150].

There is not much separation between approximation accuracy curves by values of $\bar{K}$ for the three-bit parity, though $\bar{K}$ = 3 shows slightly lower performance on the task (Figure 2.2B). For the five-bit parity, reservoirs with different $\bar{K}$ values show a clear difference in accuracy (Figure 2.3B), with $\bar{K}$ = 2 reservoirs being best. As with approximating the median function, accuracy of $\bar{K}$ = 1 reservoirs improves rapidly with L, changing from the lowest to the highest accuracy. The relative relationship between the accuracies of different $\bar{K}$-valued reservoirs that we observed is

consistent with the above-mentioned previous work, but there are some discrepancies in the actual accuracies observed. We found a much better performance of $\bar{K}$ = 1 and 3 for the three-bit parity, and $\bar{K}$ = 1 and 2 for the five-bit parity. These differences are likely due to possible differences in how the average in-degree of the reservoir is computed. In this work, the $L \times N$ edges connecting the input to the reservoir are not included in calculating $\bar{K}$, whereas they are included in the work by Snyder et al.

The recursive three-bit parity appears to be extremely difficult to approximate. All reservoirs have an accuracy ~0.5, regardless of N and L, although there is an upward trend in accuracy associated with L. Interestingly, there is no improvement with increasing N values, and it is unclear whether reservoirs with more than 500 nodes would perform any better. Given the nature of the recursive parity function, it seems that a reservoir would need to "remember" the very first bit of the input signal, which may be too difficult for any reservoir to accomplish.

2.3.4 *Estimating a Range of Functions*

To assess the flexibility of reservoir topologies, all of the possible three-bit functions were tested. For each N, L pair, we measured the mean accuracy with respect to all of the functions and all of the reservoirs - e.g., $\bar{a} = \frac{1}{100}\frac{1}{256}\sum_{i=1}^{100}\sum_{j=1}^{256} a_{ij}$ for three-bit functions. Similar to what we observed with median and parity functions, $\bar{K}$ = 3 reservoirs showed lower performance on the task (Figure 2.5).

The observations made in the three-bit space were clarified in the five-bit function space. A sample of five-bit functions showed similar trends as seen in the three-bit function set, but with greater separation between dynamical regimes (Figure 2.6). However, the ability to approximate functions is noticeably lower compared to three-bit functions. At N = 100 and L = 0.5, all prediction

accuracies are less than 0.75, regardless of $\bar{K}$. Compared to the performance in the three-bit function space where accuracy was above 0.95 using similar parameters, in the five-bit function space, it is not until the parameters are maximized that an accuracy of 0.95 is achieved.

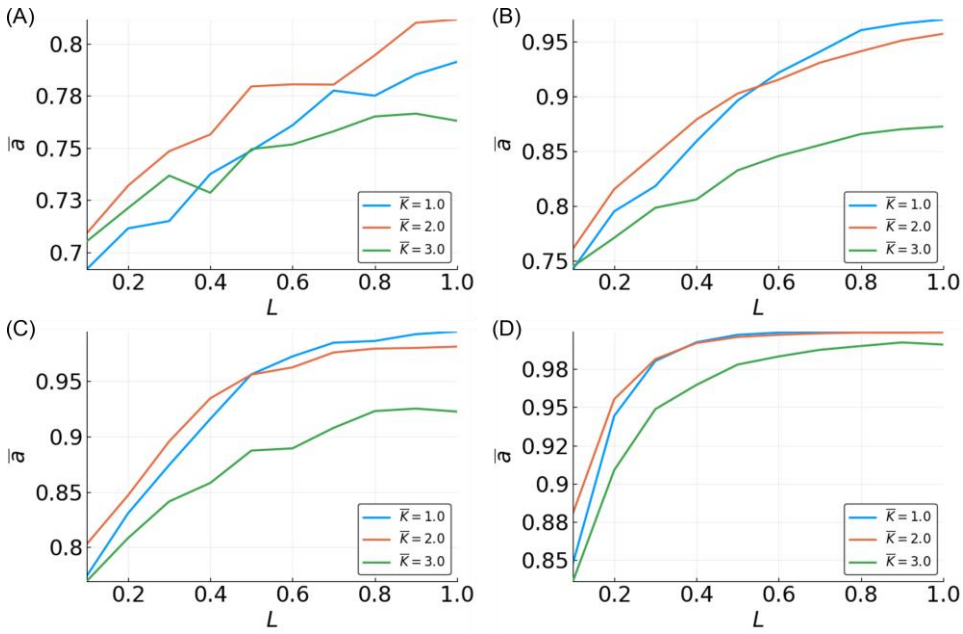


Figure 2.5. Mean accuracy, $\bar{a}$, vs. L for all three-bit functions for different $\bar{K}$-valued reservoirs. Different sizes of reservoirs are shown: N = 10 **(A)**; N = 50 **(B)**; N = 100 **(C)**; and N = 500 **(D)**.

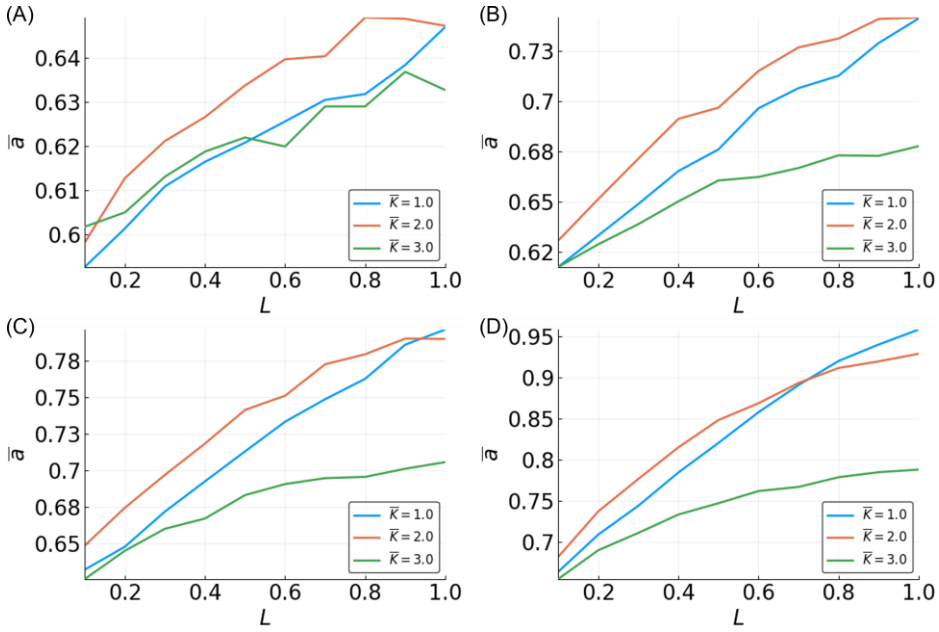


Figure 2.6. Mean accuracy, $\bar{a}$, vs. L for five-bit functions for different $\bar{K}$-valued reservoirs. Different sizes of reservoirs are shown: N = 10 **(A)**; N = 50 **(B)**; N = 100 **(C)**; and N = 500 **(D)**.

Looking at the effect of $\bar{K}$ in the five-bit function space, accuracy is increased with N and L for all $\bar{K}$, but $\bar{K}$ = 2 is most often the most accurate. The accuracy for $\bar{K}$ = 1 is higher than $\bar{K}$ = 2 after certain values of L. The value of L at which $\bar{K}$ = 1 reservoirs outperform $\bar{K}$ = 2 decreases as N increases. This effect is also partially seen in the three-bit functions (Figure 2.5) and three-bit recursive functions (Figure 2.7), although the relationship is less clear, except when looking at specific functions, such as the median and parity (Figure 2.3).

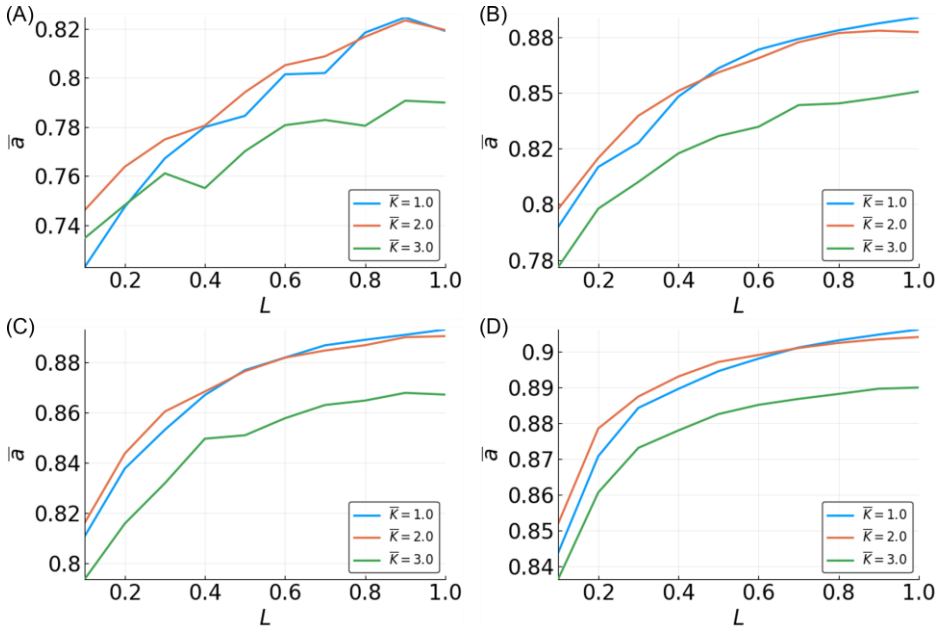


Figure 2.7. Mean accuracy, $\bar{a}$, vs. L for all recursive three-bit functions for different $\bar{K}$-valued reservoirs. Different sizes of reservoirs are shown: N = 10 **(A)**; N = 50 **(B)**; N = 100 **(C)**; and N = 500 **(D)**.

In terms of recursive functions, the accuracy is lower than both non-recursive three-bit and five-bit functions. For N = 500, L = 1, and $\bar{K}$ = 2 or less, the accuracy is above 90%. It does not appear that accuracy above ~95% is possible with the current system. This is likely due to certain functions that are essentially impossible to approximate. This is discussed below, where the temporal order of the arguments to Boolean functions is shown to have an effect.

2.3.5 *Reservoir Flexibility*

One goal of this work was to assess the flexibility of reservoirs. A highly flexible reservoir can approximate a high number of functions without any changes to the network topology or size. To assess reservoir flexibility, we compute a flexibility metric Φ_i for each reservoir i, defined as:

$$\Phi_i = median\ (a_{ij}\ - 0.5)\ /\ (mad\ (a_{ij}) + 1) \tag{2.13}$$

where the median and mad (median absolute deviation from the median) are taken over all functions g_j. Values of Φ_i are within [0, 0.5] and cannot exceed $median(a_{ij} - 0.5)$. A reservoir with a low Φ_i value cannot approximate many (or any) functions well. As the number of functions a reservoir can approximate increases, so does Φ_i.

For each combination of N and L, we have a distribution of Φ_i values, one for each of 100 BN RCs. These can be visualized as density curves (Figure 2.8). With increasing values of N and L, the distributions become increasingly right skewed and sharply peaked; meanwhile, low N, L values are characterized mostly by reservoirs with low Φ values, and high N, L values are characterized mostly by reservoirs with high Φ values. However, for many intermediary values of N and L, there is a rather flat distribution of Φ values, indicating that there is heterogeneity of flexibility in performance for reservoirs with these parameters (i.e., some are highly flexible, while some are not). The gradation of distribution shape with increasing N and L depends on the value of $\bar{K}$ (Figure S2.4A,B), although the overall trend remains. With higher values of $\bar{K}$, the distribution only shifts toward higher accuracies with higher values of L, N.

The flexibility of reservoirs approximating five-bit functions is markedly reduced compared to three-bit functions (Figure S2.5A,C,E). For all but the highest values of N and L, most of the reservoirs have only moderate flexibility (Φ = 0.25). Unlike for three-bit functions, the shape of the distributions remains the same across N and L. Consistent with the effect of $\bar{K}$ on mean accuracy, the Φ distributions for $\bar{K}$ = 1 and $\bar{K}$ = 3 are shifted to lower values as compared to $\bar{K}$ = 2.

Surprisingly, the distributions of Φ values for reservoirs approximating recursive three-bit functions behave differently than those of non-recursive three-bit functions. This was unexpected, since the mean accuracy curves (Figure 2.7) for the recursive functions resemble those for the non-

recursive functions (Figure 2.5). While Φ increases with N and L as it does for non-recursive three-bit functions, the change in the distributions does not follow the same pattern (Figure S2.5B,D,F). For low N, L values, the distribution of Φ values is wide, and becomes increasingly narrow and peaked with increasing N and L values, as compared to the narrow peaky edges and flat wide middle of the non-recursive three-bit Φ distributions.

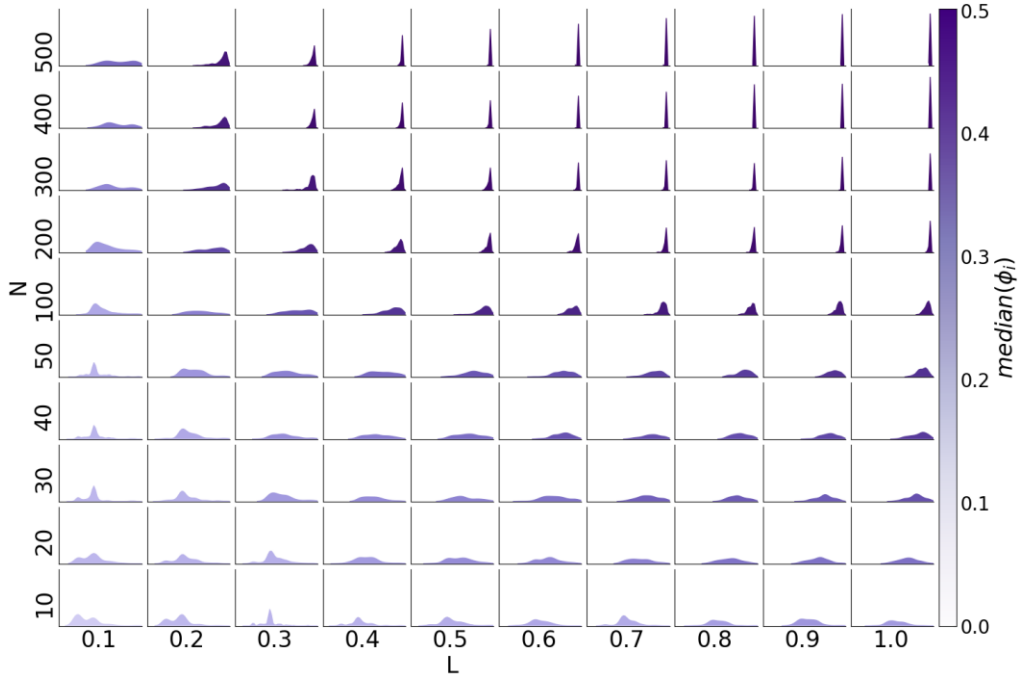


Figure 2.8. Histogram of Φ_i across all 100 reservoirs for each N, L with $\bar{K} = 2$ for three-bit functions. Each subplot represents the density for all the reservoirs with one N and L, with the x-axis being Φ, and the y-axis being number of reservoirs [0,256].

2.3.6 *Determinants of Difficulty*

To better understand why reservoirs do not perform uniformly well for all Boolean functions of a given size, we investigated possible factors related to the functions to be approximated that could

contribute to lower estimation accuracy. One possible way to compare Boolean functions with the same number of input variables is via the average sensitivity [83,155] of the function, $\bar{s}_g$. As the name suggests, this metric evaluates how sensitive the output of a function is to any change in the inputs. The average sensitivity of a function g is given by:

$$\bar{s}_g = E[s_g([u^1, u^2, \ldots, u^M])] \tag{2.14}$$

where

$$s_g([u^1, u^2, \ldots, u^M]) = \sum_{i=1}^{M} \chi[g([u^1, u^2, \ldots, u^M] \oplus e_i) \neq g([u^1, u^2, \ldots, u^M])] \tag{2.15}$$

and e_i is the unit vector with a one in the ith position and zeroes elsewhere, and $\chi[A]$ is an indicator function that gives a one if A is true and zero if A is false. The expectation is typically taken with respect to a uniform distribution over the M-dimensional hypercube.

A function that is insensitive has a low average sensitivity. For example, the constant function, $g([u^1, u^2, u^3]) = 1$, is not dependent on any of its inputs, and thus, $\bar{s}_g$ = 0. On the other hand, the three-bit parity, $g([u^1, u^2, u^3]) = u^1 \oplus u^2 \oplus u^3$, has a high average sensitivity, as it will take a different value if any of its input variables are toggled, making $\bar{s}_g = M$ = 3, the maximum possible value. The sensitivity can be interpreted as the smoothness of a function. We hypothesized that a function with high sensitivity would be more difficult to estimate, because it requires more information about the inputs, and is harder to generalize.

To investigate whether there is a relationship between average sensitivity and accuracy, we computed the average accuracy over all reservoirs, approximating all the functions with a given average sensitivity, $\bar{a}_s$ (Figure 2.9 and Figure 2.10). Here, we only discuss results for $\bar{K}$ = 2, since the results for $\bar{K}$ = 1,3 are comparable (Figure S2.6). For three-bit and five-bit functions, there is a clear inverse linear relationship between $\bar{s}_g$ and accuracy. As the average accuracy increases with

N and L, the slope of the line becomes less steep, but the relationship remains. Additionally, the approximation of functions with greater $\bar{s}_g$ does not improve as much with larger reservoirs. For low Nvalues, the relationship between sensitivity and accuracy for recursive functions closely matches that for the non-recursive functions. However, as N increases, reservoir performance for recursive functions with high sensitivity remains relatively unimproved.

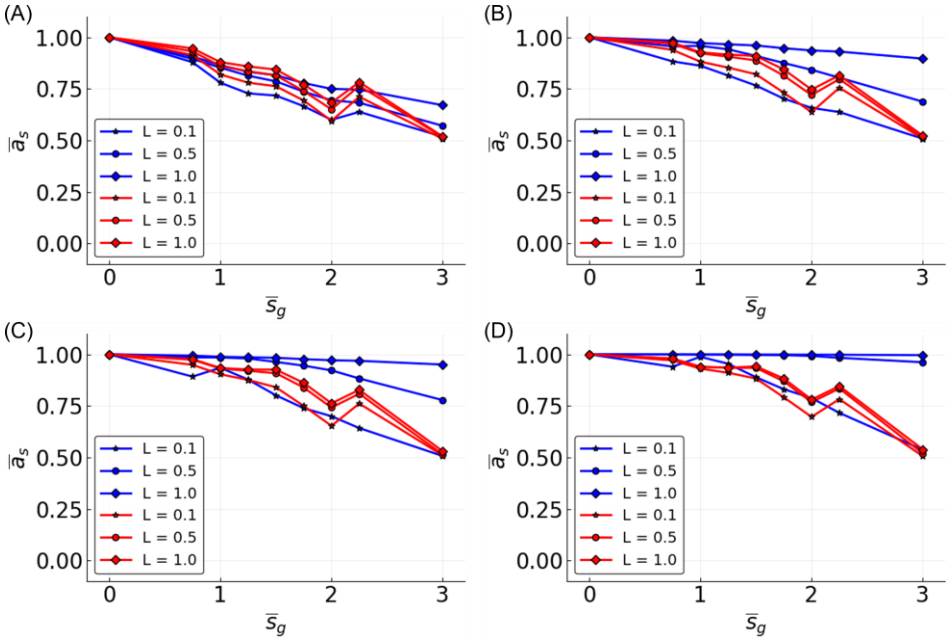


Figure 2.9. Mean accuracy, $\bar{a}_s$, vs. function average sensitivity, $\bar{s}_g$. Three-bit functions shown in blue, and recursive three-bit functions shown in red with L = 10 (stars), 50(circles), 100(diamonds). Four different values for N are shown: **(A)** N = 10; **(B)** N = 50; **(C)** N = 10; and **(D)** N = 500. Only reservoirs with $\bar{K}$ = 2 are shown. See section 2.6.Supplemental Figures for $\bar{K}$ = 1 and $\bar{K}$ = 3 (Figure S2.6).

While the relationship between the average sensitivity and accuracy appears to be mostly linear, there are clear points of deviation from linearity, most notably at $\bar{s}_g$ = 1, 2, 3, 4. Functions with the same average sensitivities can have different degrees of dependence on each variable. The

effect of each variable, u^i, on the output can be measured by the activity, α_g^i, of that variable in function g [83,156]. The activity of a variable is the probability (hence, α_g^i is between zero and one) that toggling the variable's value changes the output of the function. If a uniform distribution of input states is assumed, the sum of the activities is equal to the average sensitivity.

We examined the distribution of activities for three-bit functions and found that the points of deviation (Figure 2.9) correspond to groups of functions with the same sensitivity, but different possible activities, e.g., $\bar{s}_g = 1$, $A_g = [0, 0.5, 0.5]$ or $A_g = [1, 0, 0]$, where $A_g = [\alpha_g^1, \alpha_g^2, \alpha_g^3]$ (Figure 2.11A). For such functions, reservoir accuracy generally decreases when variables that are observed farthest in the past have higher activities. Notably, when α_g^2 and/or $\alpha_g^3 = 1$ ($\bar{s}_g = 1,2$), the function accuracy is significantly lower than for other functions with the same $\bar{s}_g$. As a result, there are large dips in the mean accuracy at integer values of $\bar{s}_g$ (Figure 2.9). Thus, the performance of the reservoir is not only dependent on the sensitivity of the function, but also the temporal distribution of activities (Figure 2.11B). For five-bit functions, the relationship between accuracy and activities is not as well defined, though still identifiable (data not shown).

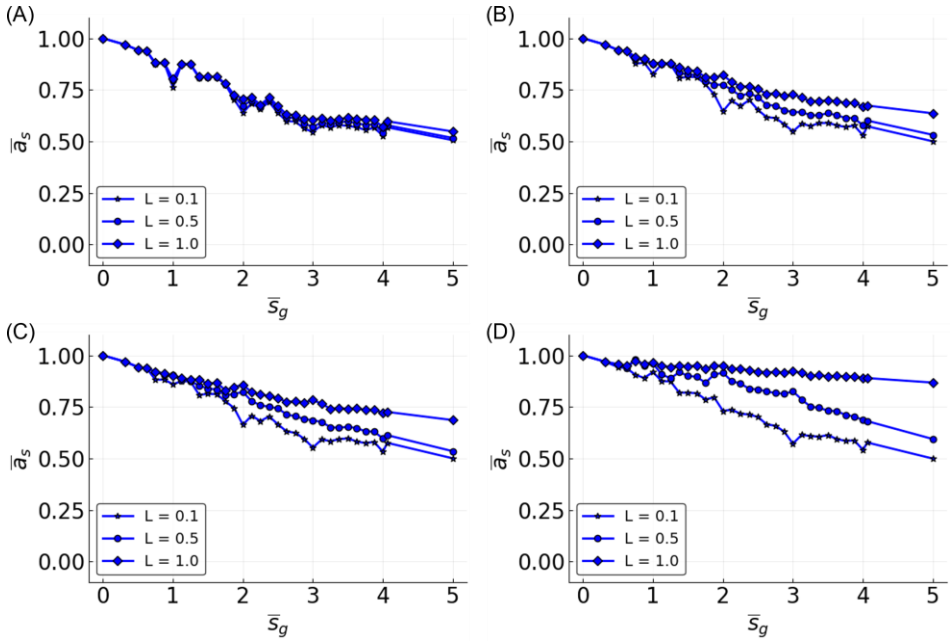


Figure 2.10. Mean accuracy, $\bar{a}_s$, vs. function average sensitivity, $\bar{s}_g$ for five-bit functions. L=0.1 (stars), 0.5 (circles), and 1 (diamonds) are given in each plot. Four different values for N are shown: **(A)** N = 10; **(B)** N = 50; **(C)** N = 100; and **(D)** N = 500. Only reservoirs with $\bar{K}$= 2 are shown.

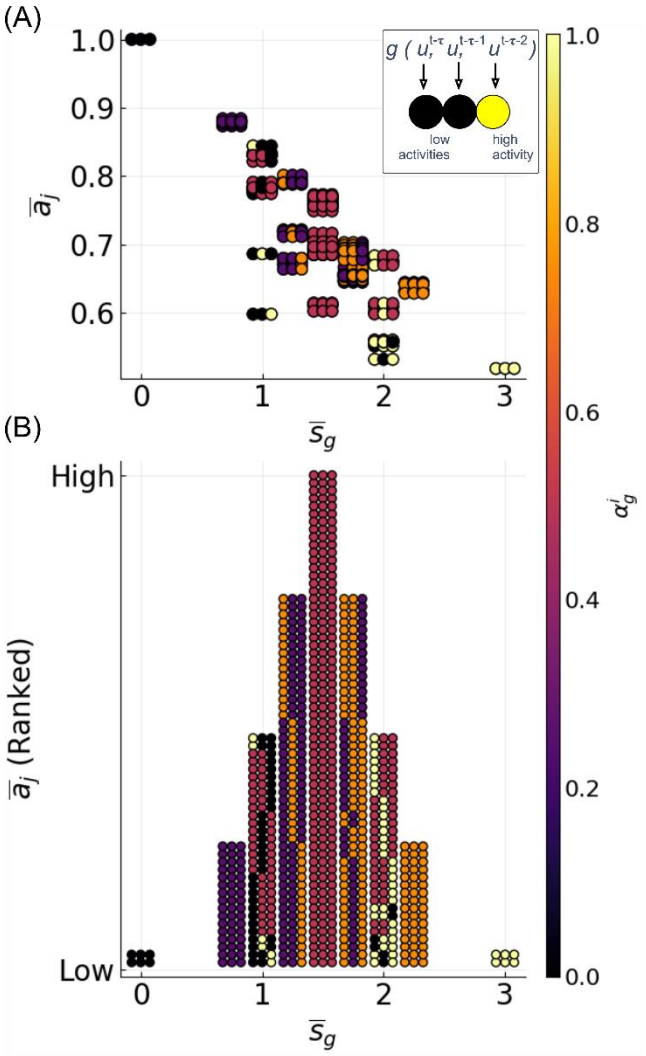


Figure 2.11. Example of mean function accuracy vs. average sensitivity with activities of each variable displayed. Data shown for three-bit functions: N = 10, L= 0.1, and $\bar{K}$ = 2. **(A)**Each of 256 functions is visualized as a horizontal triplet of circles, where each circle corresponds to a variable (left to right,$u^{t-\tau}$,$u^{t-\tau-1}$,$u^{t-\tau-2}$), colored by its activity (inset). For example, the parity function can be seen at $\bar{s}_g$ = 3, $\bar{a}_j \approx 0.5$, and A_g = [1,1,1]. **(B)** In order to more clearly see the relationship between distribution of activity and accuracy, functions are plotted by ranked accuracy rather than absolute accuracy. Here, the height of the columns here is a result of the number of functions with a given $\bar{s}_g$, and does not reflect absolute accuracy.

Recursive functions behave similarly to non-recursive functions with some exceptions (Figure S2.7A,B). Since the recursive variable is the most temporally distant, this variable having high activity causes an even greater reduction in accuracy. In fact, reservoirs cannot approximate any recursive functions with $\alpha_g^3 = 1$. There are also some cases in which reservoirs approximating functions with high activity on the recursive variable have higher accuracy compared to others with the same sensitivity. Future research could investigate whether this is related to the recursive nature of the function.

2.4 DISCUSSION

In this work, we found that the flexibility of a particular BN RC instantiation is a function of the topology and size of the reservoir, as encoded by the parameter set $(N, L, \bar{K})$. Generally, higher values of N and L result in more flexible reservoirs; however, there is heterogeneity when N or L are low, and flexible reservoirs can be found at these values. As noted in research concerning BN RCs, the optimal parameter set includes tuning the dynamics towards criticality, which leads to more flexible systems. For signal processing, a flexible reservoir will be more accurate and more efficient, requiring less searching and training when different filters are being applied to the same data. For biological systems, where survival is dependent on signal response, N, L, and $\bar{K}$ can be tuned to balance the restraints of the system with the demands of the environment.

In terms of the challenge we provided to the RCs, we found three-bit functions to be relatively easy to approximate, while five-bit functions were more difficult, essentially requiring more resources. However, three-bit recursive functions were the most difficult, where we noted some recursive functions that are essentially impossible to approximate with this system, which is likely due to a long memory requirement.

Approximating the five-bit functions depends on a large reservoir and input size, and more strongly depends on $\bar{K}$. The five-bit function space is massive, so while we were only able to sample a tiny fraction of it, we clearly saw the effect of the dynamics close to criticality on approximation accuracy. As $\bar{K}$ increased towards chaotic dynamics, the approximation became very poor, leading to large separation between dynamical regimes. Looking forward, it is likely that seven-bit or nine-bit function approximations would require increasingly more resources in terms of reservoir size and have a greater dependence on dynamics. This seems to be the case for the median and parity functions, at least [150]. We found that the relative accuracy of reservoirs with $\bar{K}$ = 1, 2, or 3 could change, depending on the value of L, which is the input connectivity into the reservoir. This is most evident with the five-bit functions in the $\bar{K}$ = 2 accuracy plot, which showed a tapering off curve (convex), where improvement with reservoir size is not linear (Figure 2.6). In the curves, as L increases, we see the $\bar{K}$ = 1 and $\bar{K}$ = 2 accuracy curves crossing each other. One possible explanation for this is that the external perturbations to the reservoir's nodes shift the dynamics of the reservoir to a less ordered regime, creating a higher fraction of perturbed nodes and resulting in a greater shift. So, the dynamics in $\bar{K}$ = 1 approach criticality with higher L values, while the dynamics in the $\bar{K}$ = 2 case moving past criticality, toward chaos, thus dropping in approximation accuracy. However, the effect of time-varying external perturbation has not been well-studied for Boolean network dynamics. Snyder et al. [150] showed that the mutual information between the reservoir and the signal increase monotonically as L increases for $\bar{K}$ = 1, so there may be other effects of increasing L at play. Although our results indicate a strong effect of mean in-degree ($\bar{K}$) in accuracy, it is possible that accuracy is also affected by other properties of reservoir topology; thus, additional in-degree distributions should be investigated.

We found that Boolean function sensitivity also plays a role in the accuracy, where more sensitive functions are more difficult to approximate. The sensitivity can be broken down further into the activities of the function variables. Functions appear to be more difficult to approximate if the activities of temporally distant variables are higher. For non-recursive functions, increasing N and L is sufficient to enable BN RCs to approximate sensitive functions; however, for non-recursive functions, large reservoirs with high values of L are not as affected. This points to a large portion of the reservoir needing perturbation to avoid problems related to function argument sensitivity.

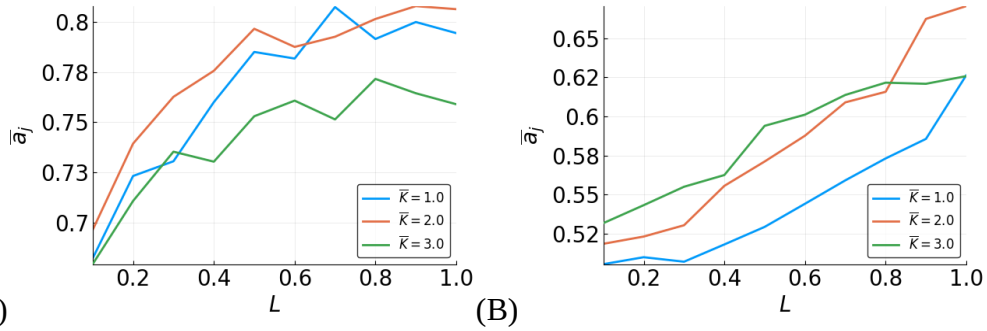
However, for recursive functions, even large reservoirs are affected by Boolean function sensitivity. This is likely due to the non-smooth output, which is difficult to approximate. Further, the recursive sliding-window filtering formulation essentially creates infinite memory, albeit fading. Many questions remain, such as what is needed for good approximations of recursive functions, and whether there is a change to the model that might help. That said, for the most part, many recursive functions are approximated very well, and that may be enough.

Overall, having an idea of the sensitivity of the processing task will inform the minimum parameter values that are necessary for a successful reservoir. This is intriguing from a biological standpoint, and it is worth exploring how signal responses that are more sensitive, especially to more historic signal values, are handled. The inability of approximating certain functions may not impinge on the use for modeling biology. When viewed through the lens of 'complex adaptive systems', recursion is a key characteristic, and one that BN RCs would need to address in order to model biological systems.

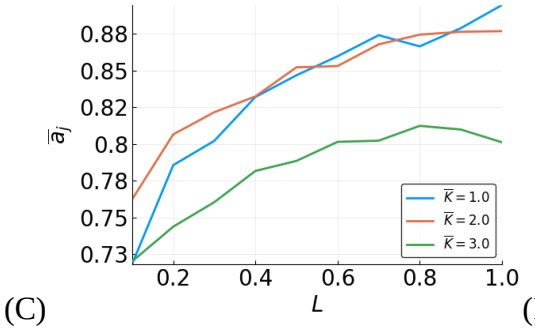
2.5 Conclusion

Reservoir computers with adequately sized Boolean network reservoirs and input sizes show that even a fixed topology is flexible enough to approximate most Boolean functions, when applied to signals in a sliding-window fashion, with high accuracy. On one hand, BN RCs have structural similarity to biological systems such as gene regulatory networks, while on the other hand, they are useful objects for operating on signals, such as through the application of recursive filters to biological data. We observe that Boolean networks that are closer to the critical dynamics regime lead to more flexible reservoirs. Recursive functions are found to be the most difficult to approximate, owing to the necessity for approximating a function with hidden variables. Although most recursive functions can be approximated, we find that some can never be approximated. Boolean functions with more arguments are more strongly dependent on the dynamics of the system for accuracy, leading to a greater separation in accuracy curves that correspond to different dynamical regimes. We find a connection between Boolean function sensitivity, an estimate of the smoothness of a function, and the ability to approximate a given function.

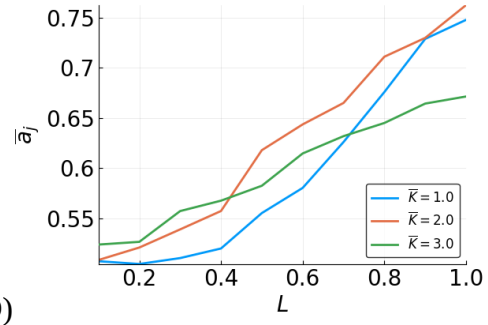
2.6 Supplemental Figures



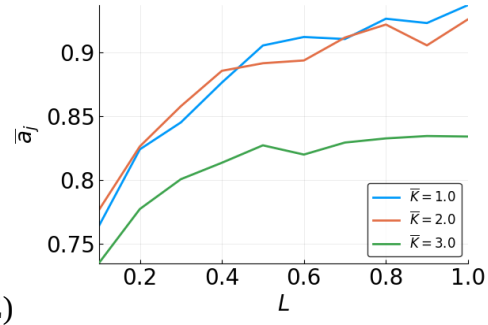
(A) (B)


0.88
0.85
0.82
0.8
0.78
0.75
0.73
$\overline{a}_j$
0.2
0.4
0.6
0.8
1.0
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

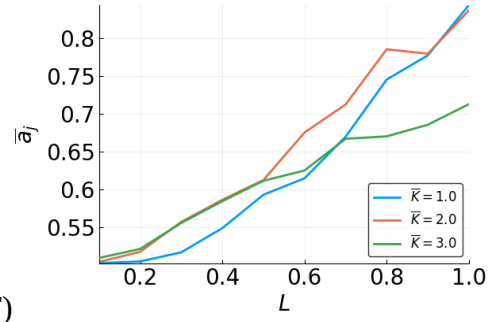
(C)


0.75
0.7
0.65
0.6
0.55
$\overline{a}_j$
0.2
0.4
0.6
0.8
1.0
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

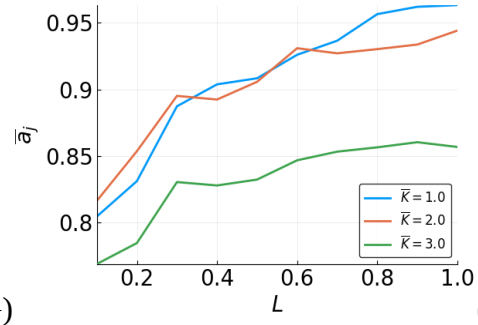
(D)


0.9
0.85
0.8
0.75
$\overline{a}_j$
0.2
0.4
0.6
0.8
1.0
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

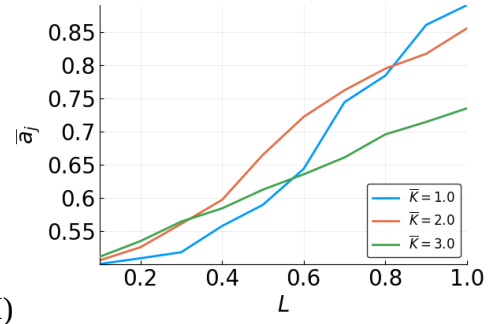
(E)


0.8
0.75
0.7
0.65
0.6
0.55
$\overline{a}_j$
0.2
0.4
0.6
0.8
1.0
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

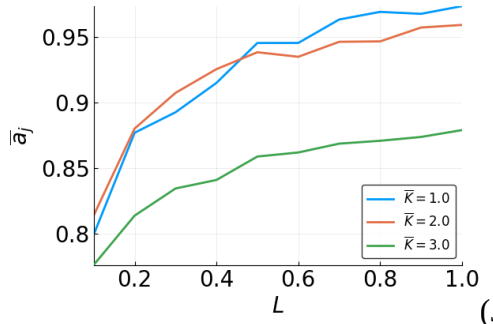
(F)


0.95
0.9
0.85
0.8
$\overline{a}_j$
0.2
0.4
0.6
0.8
1.0
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

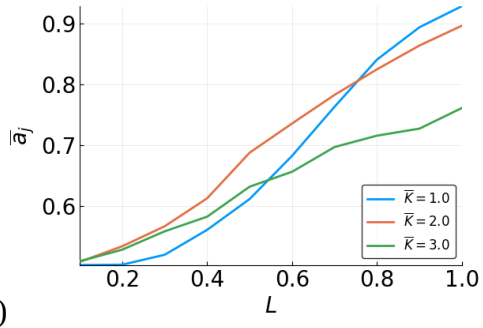
(G)


0.85
0.8
0.75
0.7
0.65
0.6
0.55
$\overline{a}_j$
0.2
0.4
0.6
0.8
1.0
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

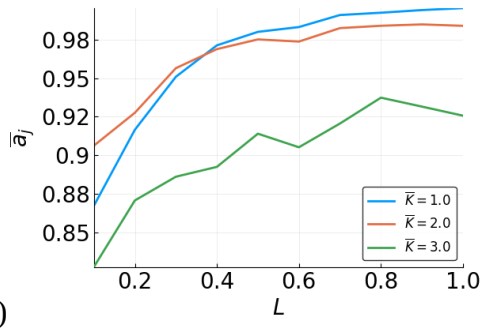
(H)


0.95
0.9
0.85
0.8
$\overline{a}_j$
0.2
0.4
0.6
0.8
1.0
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

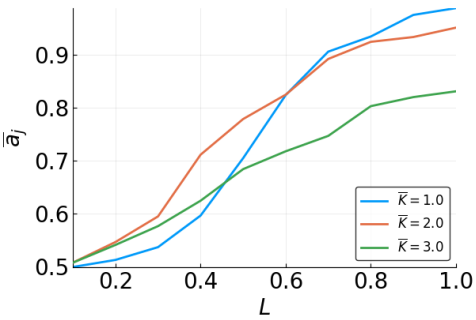
(I)


0.9
0.8
0.7
0.6
$\overline{a}_j$
0.2
0.4
0.6
0.8
1.0
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

(J)


0.98
0.95
0.92
0.9
0.88
0.85
$\overline{a}_j$
0.2
0.4
0.6
0.8
1.0
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

(K)


0.9
0.8
0.7
0.6
0.5
$\overline{a}_j$
0.2
0.4
0.6
0.8
1.0
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

(L)

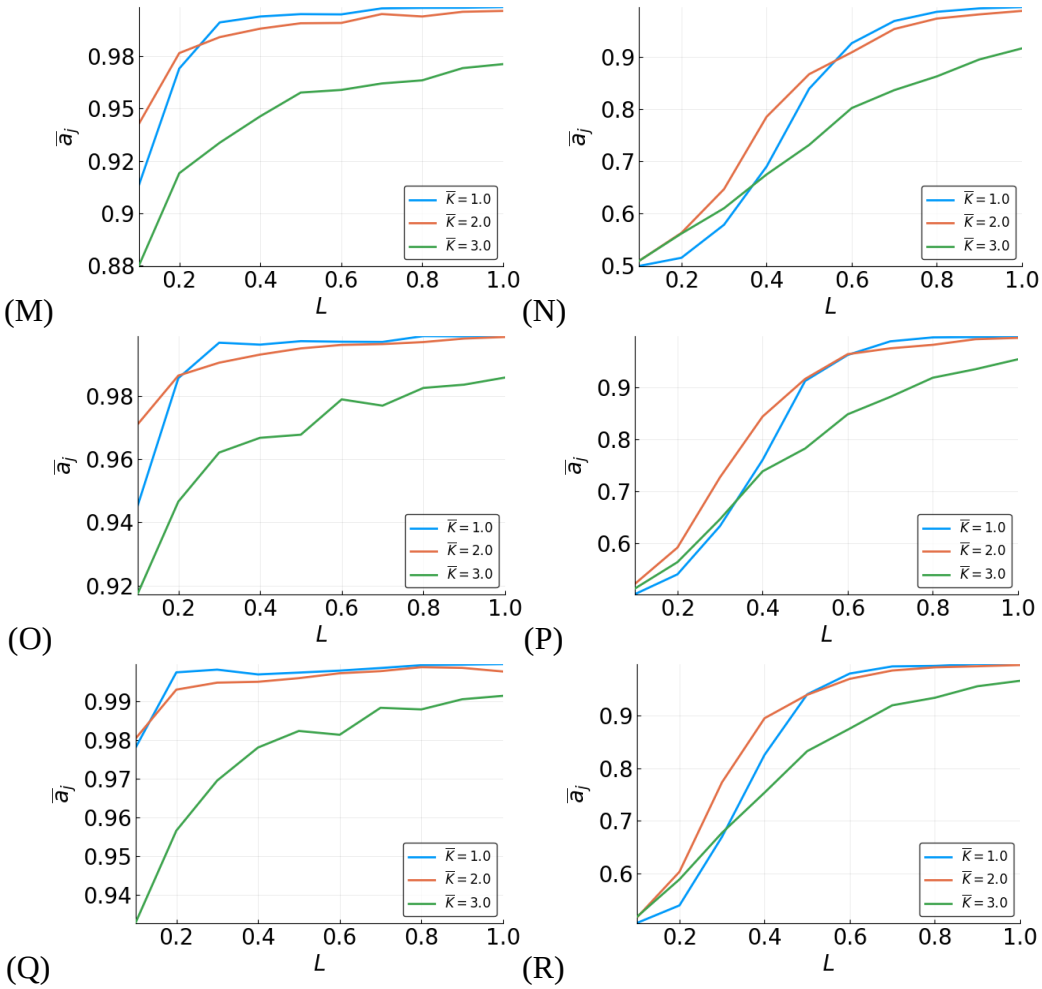
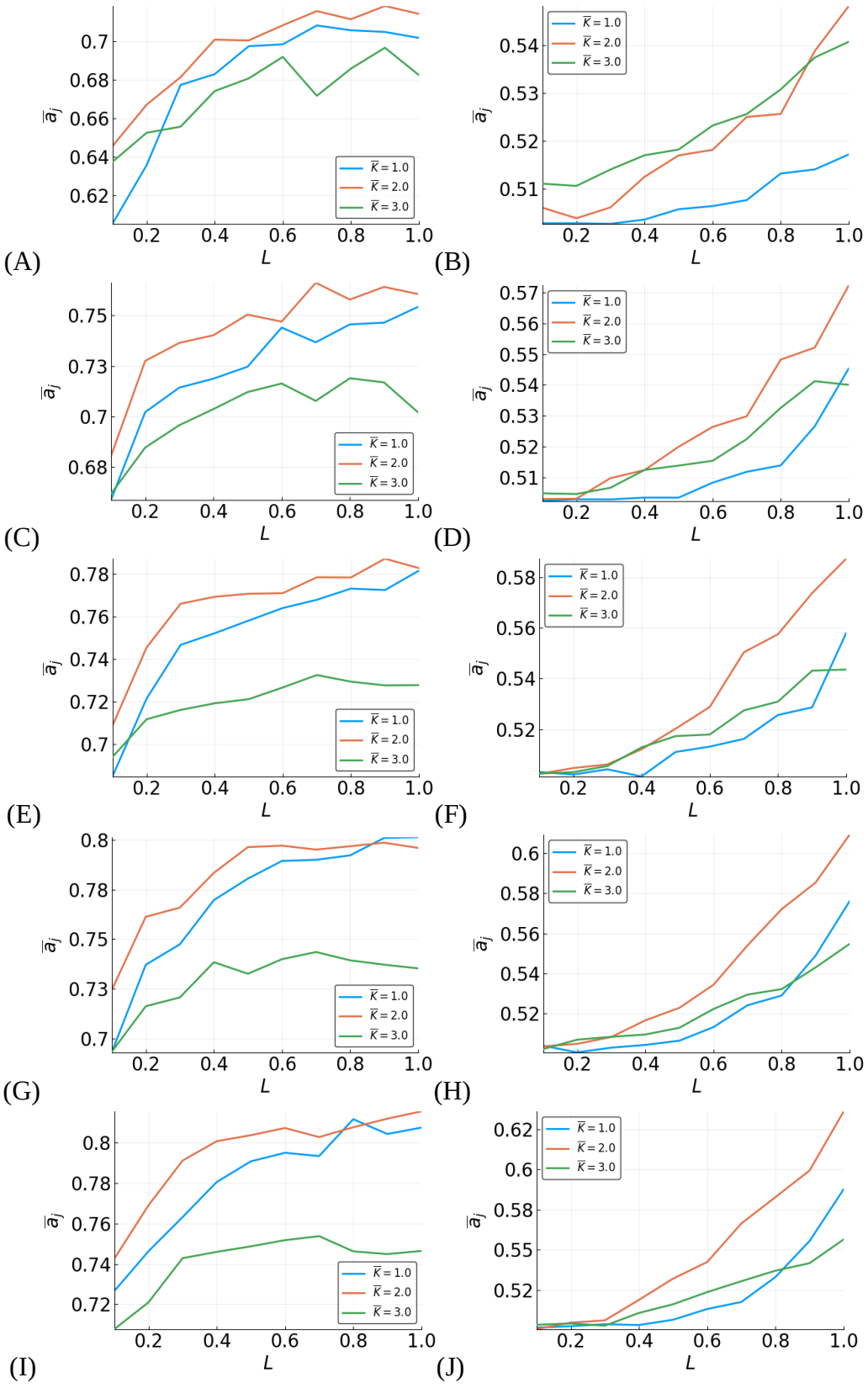


Figure S2.1. Mean accuracy, $\bar{a}_j$ vs L for the three-bit median (Left Column) and parity (Right Column) functions for different $\bar{K}$-valued reservoirs with N= 10 **(A,B)**, 20 **(C,D)**, 30 **(E,F)**, 40 **(G,H)**, 50 **(I,J)**, 100 **(K,L)**, 200 **(M,N)**, 300 **(O,P)**, 400 **(Q,R)**.


(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
(J)
$\overline{a}_j$
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

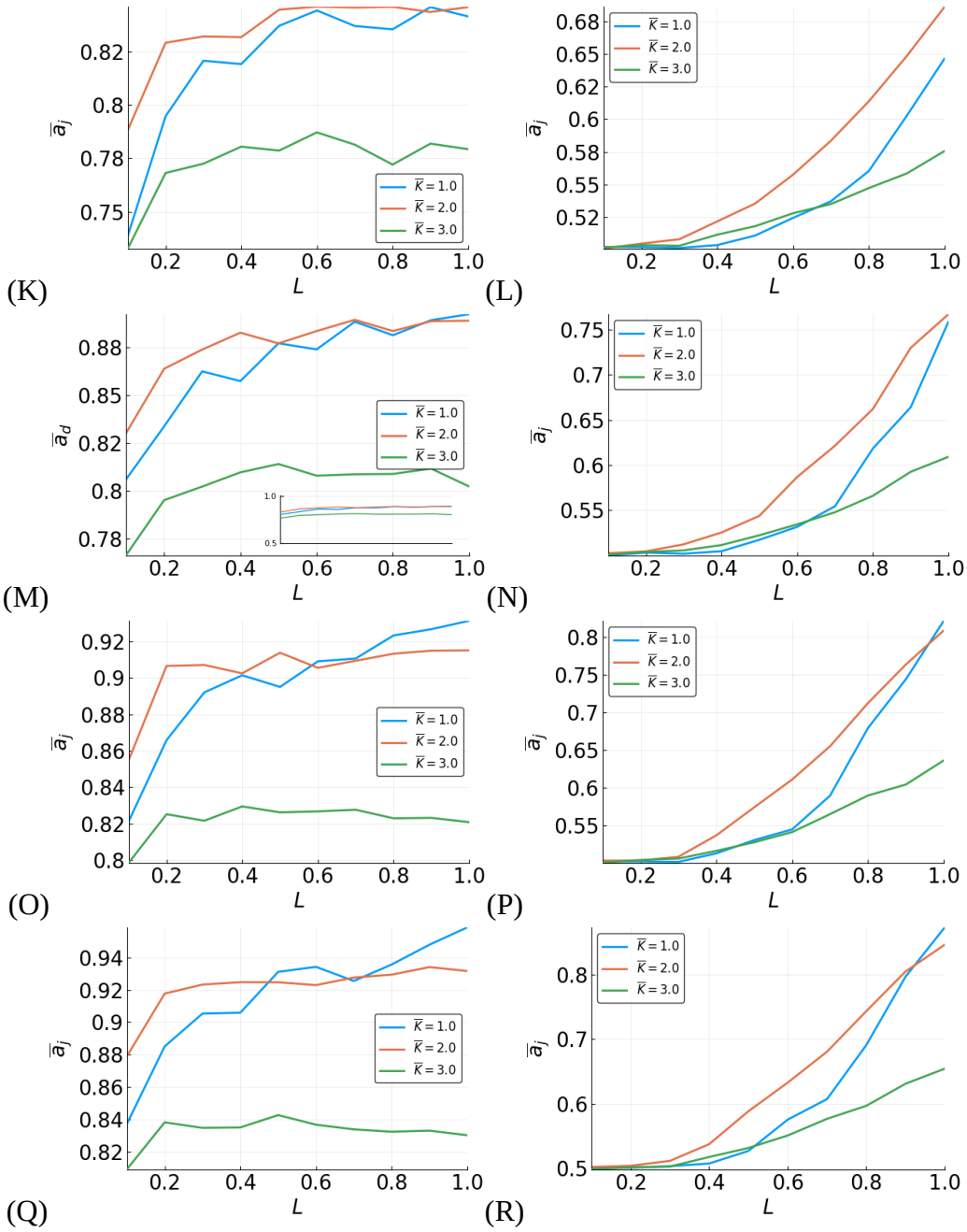
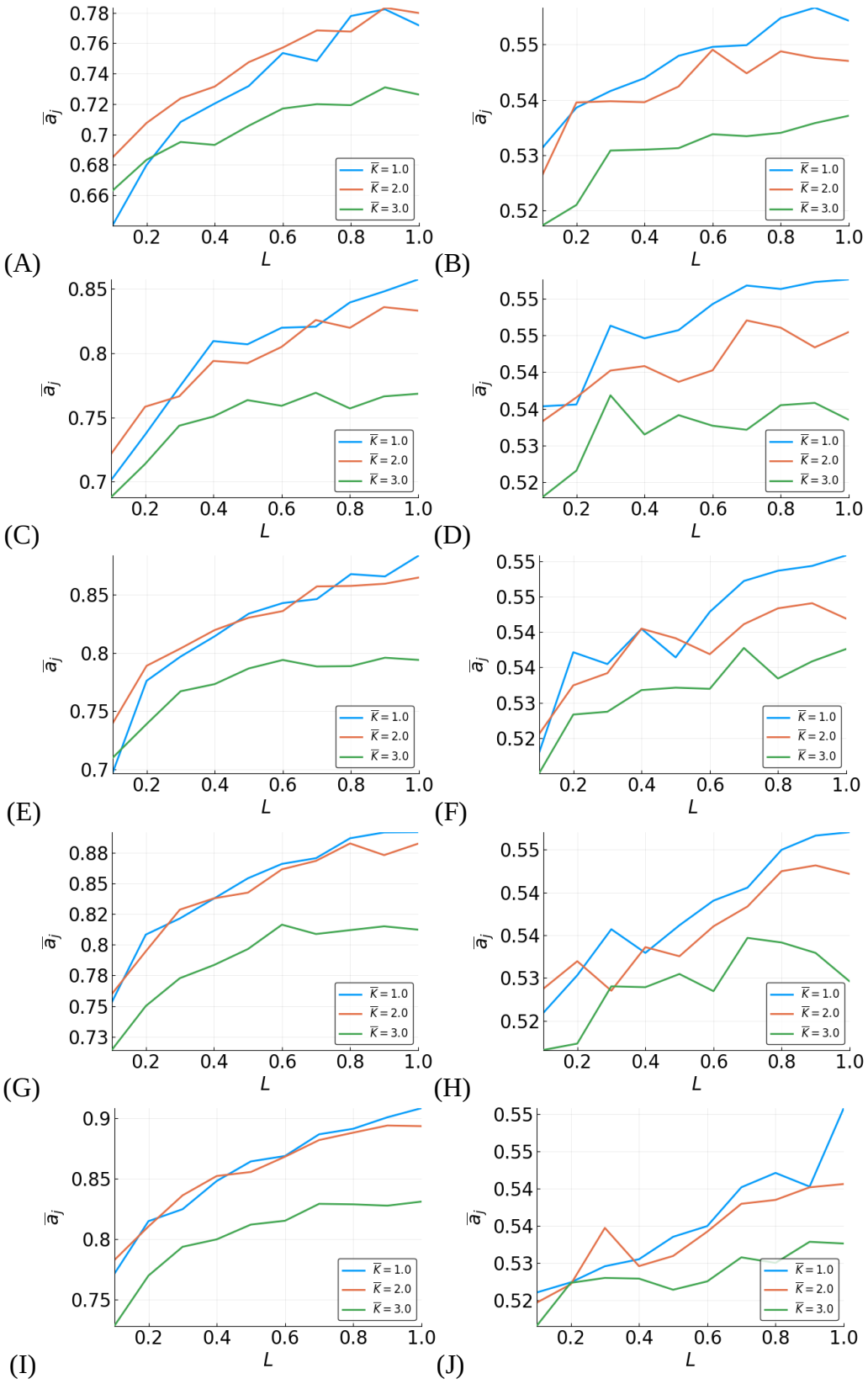


Figure S2.2. Mean accuracy, $\bar{a}_j$, vs L for the five-bit median (Left Column) and parity (Right Column) functions for different $\bar{K}$-valued reservoirs with N= 10 **(A,B)**, 20 **(C,D)**, 30 **(E,F)**, 40 **(G,H)**, 50 **(I,J)**, 100 **(K,L)**, 200 **(M,N)**, 300 **(O,P)**, 400 **(Q,R)**.


(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
(J)
$\overline{a}_j$
L
$\overline{K} = 1.0$
$\overline{K} = 2.0$
$\overline{K} = 3.0$

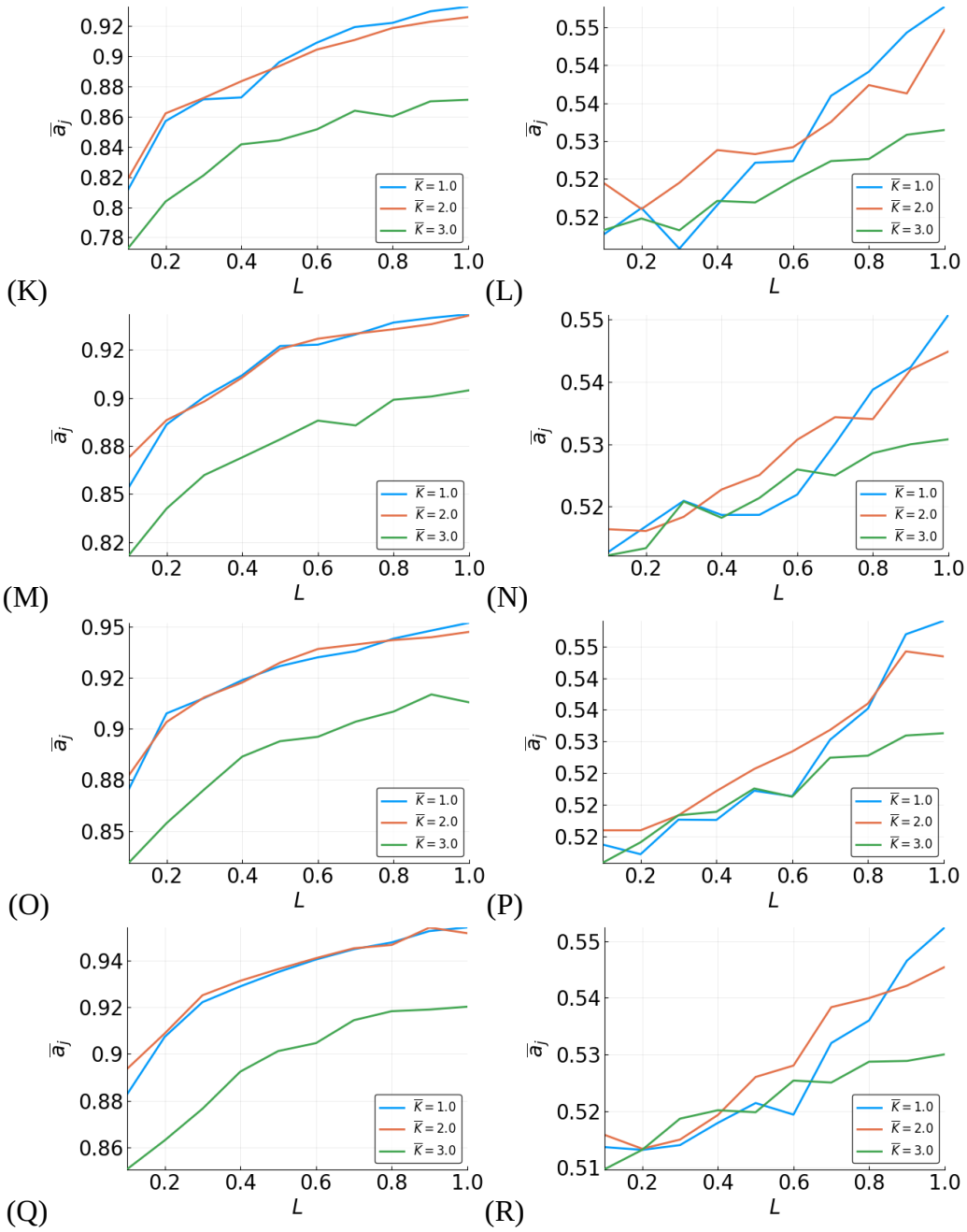


Figure S2.3. Mean accuracy, $\bar{a}_j$, vs. L for the three-bit recursive median (Left Column) and parity (Right Column) functions for different $\bar{K}$-valued reservoirs with N= 10 **(A,B)**, 20 **(C,D)**, 30 **(E,F)**, 40 **(G,H)**, 50 **(I,J)**, 100 **(K,L)**, 200 **(M,N)**, 300 **(O,P)**, 400 **(Q,R)**.

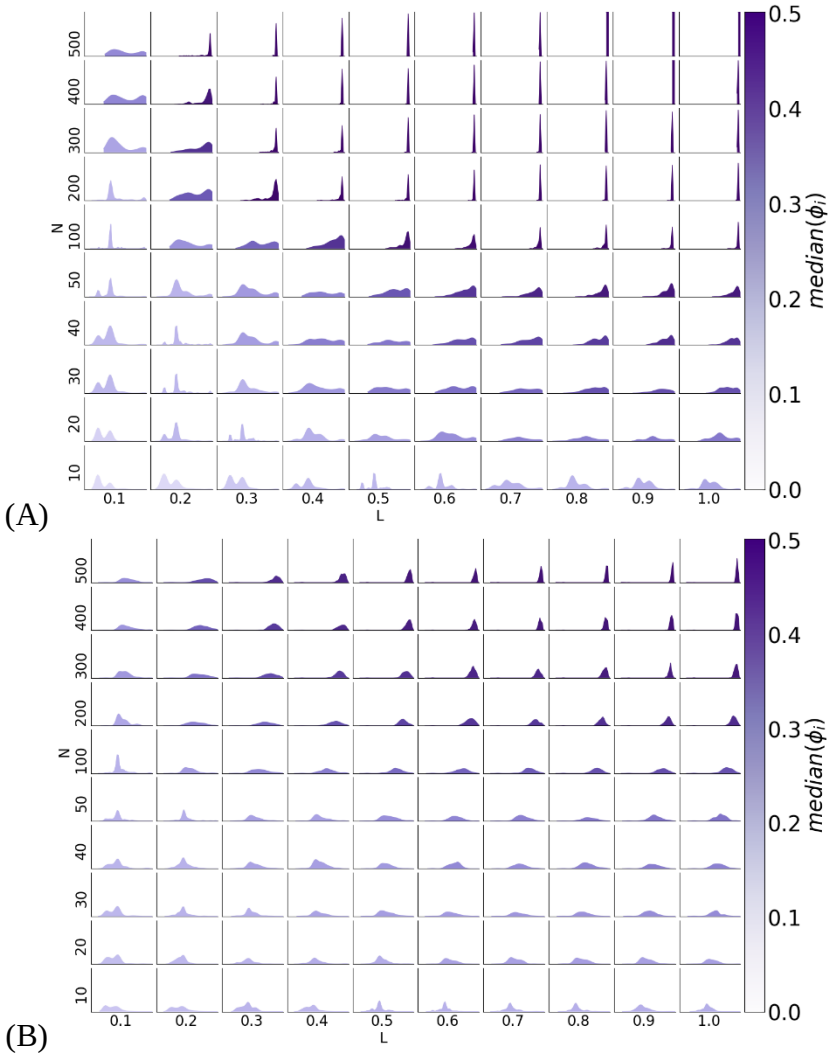


Figure S2.4. Histogram of Φ_i across all 100 reservoirs for each N, L with $\bar{K}$ = 1 **(A)**, $\bar{K}$ = 3 **(B)** for three-bit functions. Each subplot represents the density for all the reservoirs with one N and L, with the x-axis being Φ and the y-axis being number of reservoirs [0,256].

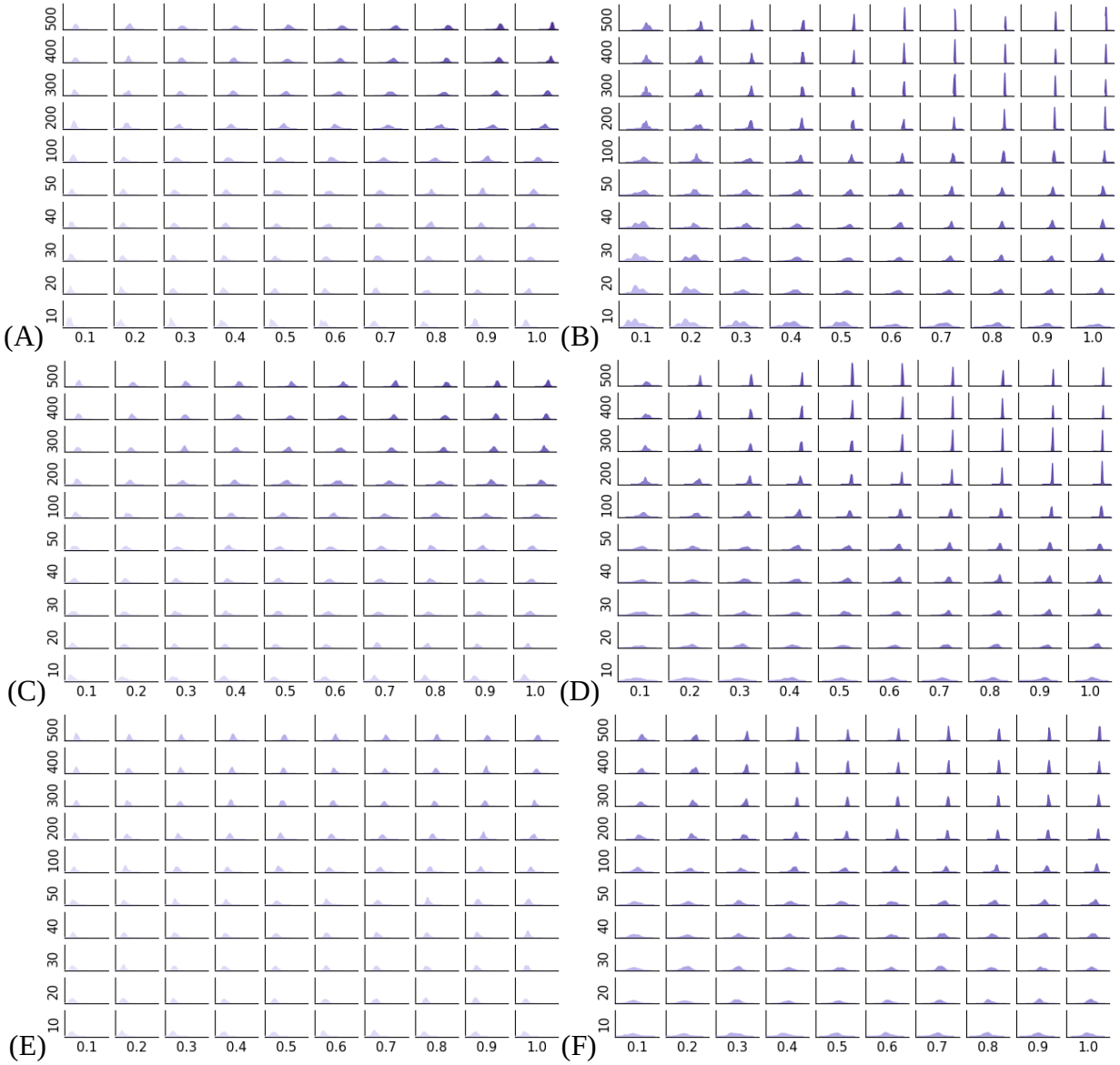


Figure S2.5. Histogram of Φ_i across all 100 reservoirs for each N, L for five-bit functions (Left column) and recursive three-bit functions (Right column). Reservoirs with the three values of $\bar{K}$ are shown: $\bar{K} = 1$ **(A,B)**, $\bar{K} = 2$ **(C,D)**, and $\bar{K} = 3$ **(E,F)**. Each subplot represents the density for all the reservoirs with one N and L, with the x-axis being Φ and the y-axis being number of reservoirs.

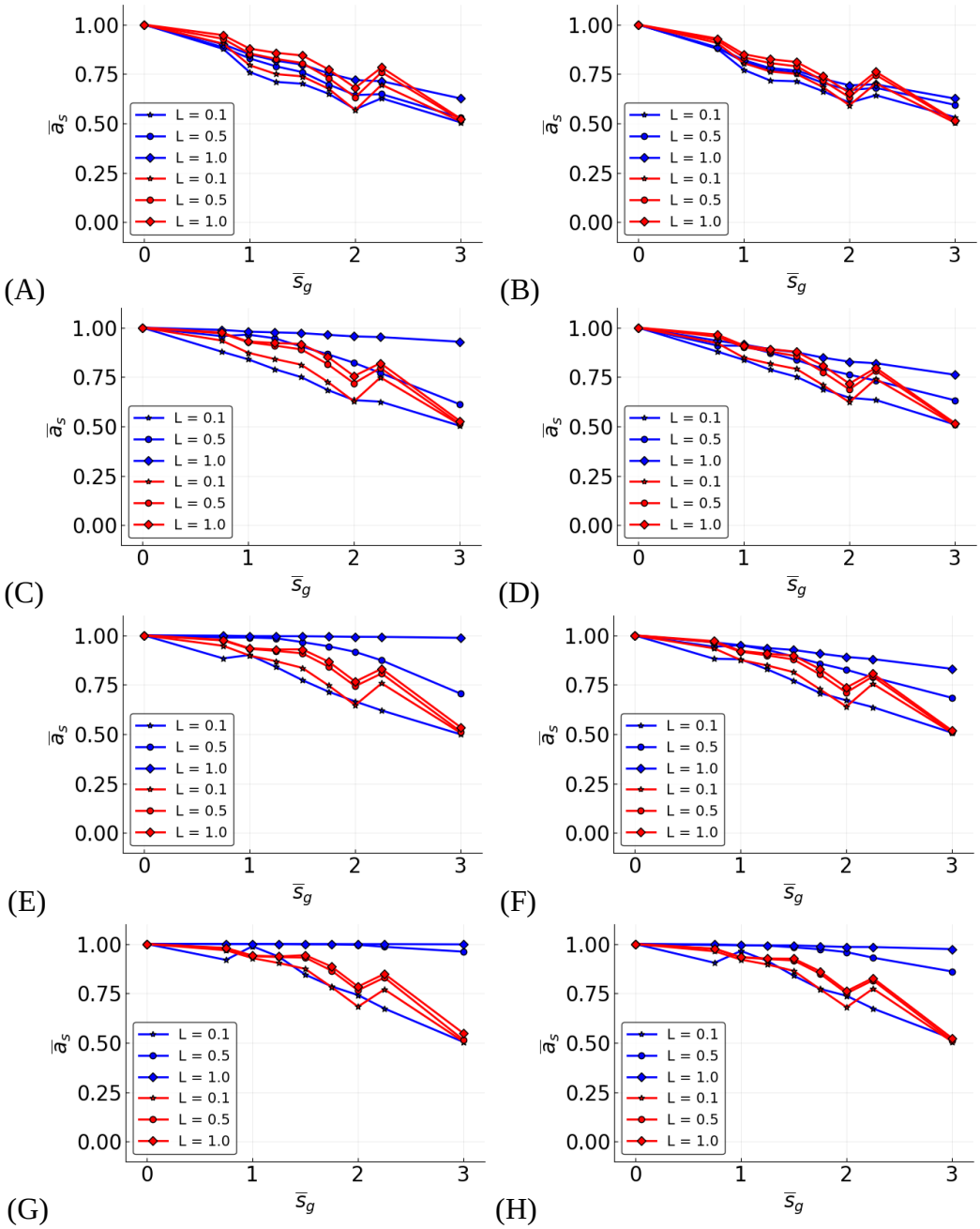
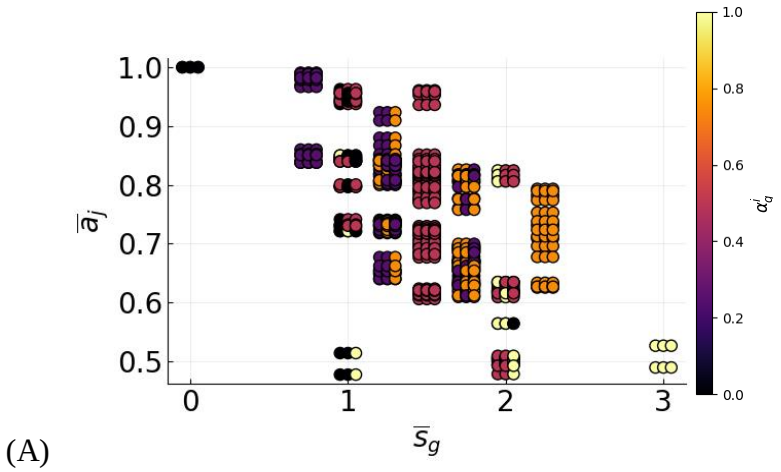
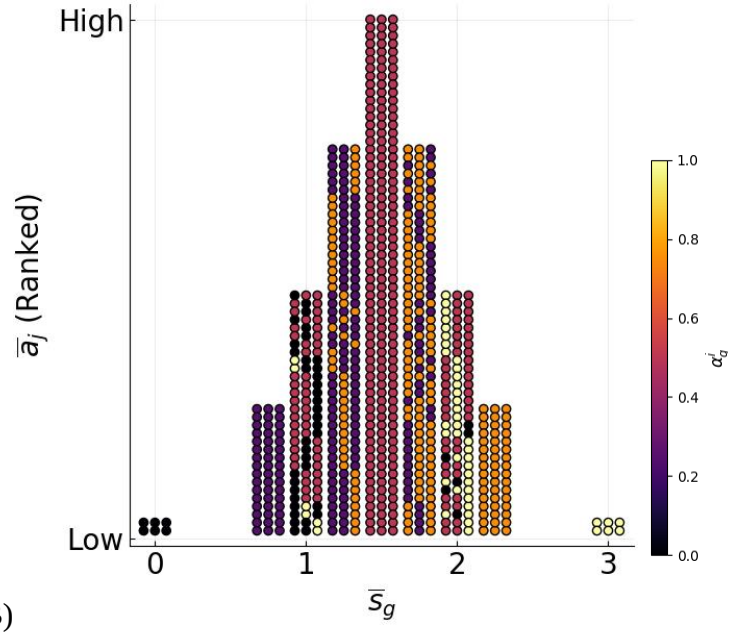


Figure S2.6. Mean accuracy, $\bar{a}_s$, vs. function average sensitivity, $\bar{s}_g$. three-bit functions shown in blue and recursive three-bit functions shown in red with L = 0.1 (stars), 0.5(circles), 1(diamonds). Each row shows data from reservoirs with different values of N, from top to bottom N= 10 **(A,B)**, N = 50 **(C,D)**, N = 100 **(E,F)**, N = 500 **(G,H)**. Columns show $\bar{K}$ = 1 (Left) and $\bar{K}$ = 3 (Right).



(A)



(B)

Figure S2.7. Example of mean function accuracy vs. average sensitivity with activities of each variable displayed. Data shown is for recursive three-bit functions: N = 10, L= 0.1, and $\bar{K}$ = 2. **(A)** Each function is visualized as a horizontal triplet of circles, where each circle corresponds to a variable (left to right,$u^{t-\tau}$,$u^{t-\tau-1}$,y^{t-1}), colored by its activity. **(B)** In order to more clearly see the relationship between distribution of activity and accuracy, functions are plotted by ranked accuracy rather than absolute accuracy.

Chapter 3. CHARACTERIZING THE IMPACT OF COMMUNICATION ON CELLULAR AND COLLECTIVE BEHAVIOR

3.1 INTRODUCTION

Communication between cells is a key property of living systems, from unicellular microbes to multicellular organisms. By signaling to each other, cells are able to coordinate their activity, thereby attaining structural and functional complexity. Fundamentally, these complexities are achieved by regulating which cellular behaviors occur within a population and how those behaviors are organized. In practice, organisms have evolved communication systems that are diverse in both mechanism and function.

In microbial populations, cell-cell communication systems are pervasive and utilized for a variety of population functions. For example, signaling is used to propagate information through a population in order to provide long range coordination, as in the use of ion channels to regulate metabolic states during growth cycles in biofilms [101,157]. It can be used to aggregate environmental information in order to execute a singular collective behavior, such as infectious bacteria like *V. cholera* and *P. aeruginosa* relying on quorum sensing to collectively express virulence factors when pathogenicity will be maximized [158,159]; motile bacteria like *E. coli* secreting a chemical trail to collectively migrating toward food sources [160]; and sporulating bacteria like *B. subtilis* collectively sporulating in response to quorum signals when there is not enough food for colony survival [161]. Cell-cell signaling can also be used to manage subpopulations with different, often mutually exclusive, functions that need to act in concert, such as in the generation of matrix- and

surfactant-producing cells in *B. subtilis* colonies during colony expansion [162,163]; and the maintenance of regularly spaced nitrogen-fixing heterocysts and photosynthetic vegetative cells in filamentous cyanobacteria for survival in low-nitrogen conditions [164,165]. Further examples of microbial communication systems can be seen in colony morphogenesis or multispecies biofilm maintenance [166,168].

In multicellular organisms, cell-cell communication is so vital that it is recognized as a defining feature of multicellularity [169]. Indeed, it has been proposed that an expansion in the genes dedicated to cell-cell communication accompanied the transition from unicellular to multicellular life [170]. As in microbial systems, cell-cell signaling in multicellular organisms can serve to synchronize cells in a common behavior or manage diverse subpopulations; however, there is much more breadth and complexity in multicellular communication systems. From simple to complex organisms, across evolutionary kingdoms, cell-cell communication is heavily involved in nearly all multicellular biological operations. For example, during development multicellular organisms rely on cell-cell communication cues for functions like temporal and spatial cellular differentiation (plant stem morphogenesis [171], nematode neuronal fate decisions [172]); directional cell migration (zebrafish organogenesis [173], mammalian vessel formation [174]); and controlled cell death (avian limb development [175], insect body segmentation [176]). In mature organisms, homeostasis is maintained by cell-cell communication through functions like population size control (murine T cells [177] and macrophage/fibroblasts [178] populations), maintaining multicellular structures (animal corneal maintenance [179] and bone remodeling [180]), and coordination between subsystems (human wound healing [181]). Mature organisms must also navigate their environment as a unit, using cell-cell communication to perform complex functions (mollusc neuromuscular coordination [182], vertebrate adaptive immunity [183]) necessary for survival.

In addition to naturally occurring systems, cell-cell communication is increasingly being utilized as a tool in synthetic systems. Many of the communication mechanisms used in living systems can be adapted for synthetic systems. Current research areas include the synthesis of populations capable of swarming behavior [184], reliable spatial patterning [185], organizing into multicellular structures [186] ,biological computing [187], and complex population functions like biofuel production [188]. For a thorough review of synthetic cell-cell communications in both prokaryotes and eukaryotes, see Henning et al.(2015) [189].

While research has elucidated the form and function of cell-cell communication in many biological systems across both unicellular and multicellular organisms, there are often other phenomena at play that confound the effects of communication. For example, the architecture of the genetic network, stochasticity in gene expression, asymmetric cell division, and cell cycle heterogeneity can all contribute to behavioral heterogeneity within a population [190-193]. In addition to these factors, mechanical interactions, cell death and birth, and environmental conditions can influence the organization of cell types [194-197]. Furthermore our knowledge of how communication can impact system behavior is limited by which systems have evolved, which is driven by the functional requirements and historical contingency of each specific system.

Given that cell-cell communication is such a vital aspect across many diverse systems, understanding the extent of its effects can provide widely-applicable insights. Thus, with the work reported here we strive to push forward the general understanding of what impact cell-cell communication can have on cellular and population behavior. We can apply this understanding to learning how current systems might have evolved, including the transition from unicellular to multicellular life. We can also more fully understand how extant systems operate and how modifications to communication can affect population dynamics. Similarly, such knowledge is

useful in understanding the role of communication in disease-altered states and how disease effects might be remediated by cell-cell communication-based intervention. In addition, for synthetic systems, it is necessary to understand the role of communication so that systems can be designed accurately and efficiently, and undesirable collateral effects can be avoided.

In order to isolate the effects of communication so that we may understand the general principles relating cell-cell communication to cellular and collective behavior, we use an abstract *in silico* model of cellular populations. Specifically, we use a 3D multiscale cellular population model, with dynamic intracellular networks interacting via diffusible signals to form intercellular connections (Figure 3.1). In this way we can have a more detailed understanding of the impact of communication at both the cell and population levels. Importantly, cell-cell communication is not a simple present/absent phenomenon in living systems. The mechanism, architecture, and strength of communication are all variable and likely have functional consequences. Therefore, in our work we focus on two key signaling parameters: the effective interaction distance, which is the distance at which cells are able to interact via signals; and receptor activation threshold, which is how much signal a cell must receive before its receptor is activated. Both of these parameters have previously been shown to tune cell's relative response to self-generated versus non-self-generated signal [103,198]. Thus, by varying these parameters, which are variable between and within biological systems, we can vary the degree of communication and more finely explore the effects of communication.

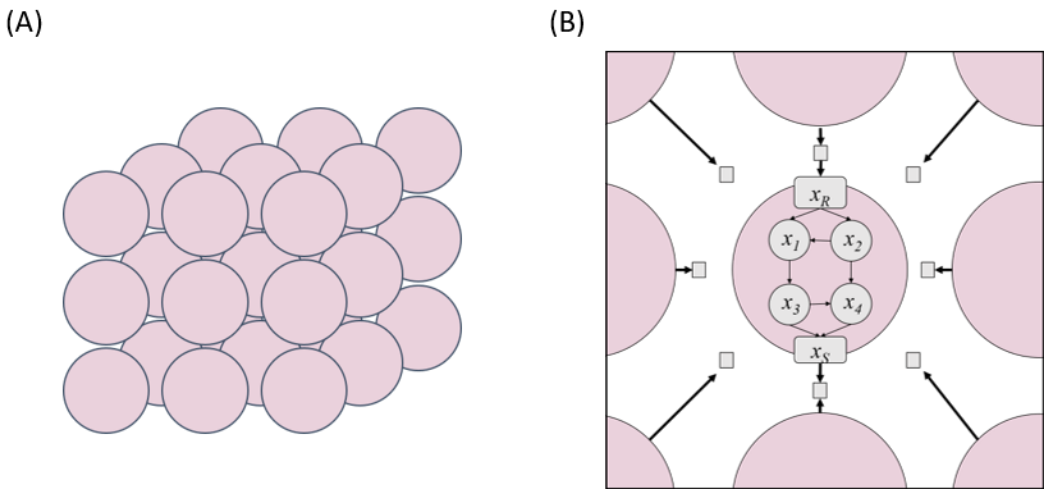


Figure 3.1. 3D structured population of communicating cells. **(A)** The cellular population model consists of cells (pink circles) that are arranged in three dimensions on a lattice-structured grid. **(B)** Cells are described by internal networks and communicate with each other by secreting and sensing diffusible molecules (gray squares).

3.2 METHODS

3.2.1 *Mathematical Model*

In order to investigate the impact of communication on cellular and population behavior, we take an *in silico* modeling approach to study a generalized population of communicating cells. Briefly, our framework consists of a population of individual cells, organized on a 3D grid, which communicate through diffusible molecules. Each cell is described by an internal Boolean network, consisting of binary-valued regulatory nodes and signaling nodes. While regulatory nodes largely determine the behavior of the cell, signaling nodes are responsible for communication between cells. By iterating the state of the system over time, we simulate the dynamics of both cells and population. For simplicity, we will refer to a population of cells with identical cellular BNs as a tissue for the rest of the paper. Since a tissue may have different signaling parameters, we will refer to a tissue with specific signaling parameters as a tissue sample. By simulating tissues with

different signaling parameters, we can probe the effect of communication on cellular and tissue dynamics.

3.2.1.1 Cells as Boolean Networks

Each cell in our simulation framework is modeled as a Boolean network (BN). BNs have a long history of being used as biological models, especially as gene regulatory networks [42]. They are simple to implement and analyze, but still capable of recapitulating many complex behaviors observed in biological systems [44,145]. Specifically, the BN of cell i is a network defined on a set of n binary-valued variables,

$$X_i = \{x_1^i, \ldots, x_n^i\}, \tag{3. 1}$$

where $x_j^i \in \{0, 1\}$ represents the expression or activation of node j. The nodes in the cellular networks are divided into two categories: regulatory nodes and signaling nodes. Signaling nodes come in pairs of a signal node, which secretes a diffusible molecule, and a receptor node, which is activated by the (sufficient) presence of that molecule.

Over time the state of the cell, $X_i(t)$, changes via updating the values of the regulatory and signaling nodes. The value of each regulatory node at time $t + 1$is determined by the values of k input nodes at time t by means of a corresponding Boolean update function, $f_j^i: \{0, 1\}^k \to \{0, 1\}$. Thus, the Boolean value of each regulatory node $x^i{}_j$ is given by:

$$x_j^i(t + 1) \; = \; f_j^i(x_{j_1}^i(t), x_{j_2}^i(t), \ldots, x_{j_k}^i(t)). \tag{3. 2}$$

The function f_j^i can also be described by the table of output values (truth table) it gives for each combination of argument values. There are many schemes for generating the architecture of the wiring between nodes (wiring diagram) and truth tables for Boolean networks. For simplicity, we

follow the original formulation of the random Boolean network (RBN) scheme [37]. For the wiring diagram, k is constant for all genes; the k input genes for each regulatory gene are chosen randomly from the regulatory and receptor genes. The signal node is never used as an input in order to keep communication distinct from internal regulation. For the truth tables, elements are assigned values of 0 or 1 with a bias $p = P\{f_j^i = 1\}, j = 1, \ldots, n$, which is the probability that the function takes on the value of 1for a combination of input values. In closed BNs (i.e., with no external inputs), k and p can be varied to tune the network's response to perturbations, ranging between ordered dynamics (insensitive to perturbations) and chaotic dynamics (highly sensitive to perturbations).

The two types of signaling nodes, signals and receptors, each behave differently. The signal node follows the same rules as the regulatory nodes. The receptor node updates using a threshold function on the local concentration of its corresponding signal - described in section 3.2.1.3.

3.2.1.2 Tissue Architecture

In our simulations, each of the N cells in a tissue is described by the same BN (identical wiring diagram and truth tables) and occupies a grid point on a 3D structured lattice. Specifically, cells are located in the center of voxels, which have a volume of $V_c = H^3$ where H is the lattice size. The lattice has periodic boundary conditions on all three axes, such that all cells are signaling identically with the same neighborhood structure. This allows for observing patterns that might appear in much larger tissues while only simulating a small tissue. It should be noted that simulated tissues are physically static; while the states of the cells change over time, cells do not die, undergo mitosis, or migrate.

3.2.1.3 Cell-cell Communication via Diffusible Molecules

We include signaling between cells by modeling the secretion and sensing of a single diffusible molecule, m. A cell i releases molecule m with a secretion rate of $\eta(x_S^i)$ molecules per second, which depends on the Boolean state of its designated signal node x_S^i. We assume that $\eta(0) = 1$and $\eta(1) = \alpha,\ \alpha > 1$, to account for basal and active expression, respectively. We make this assumption with no loss of generality since it is equivalent to normalizing active expression by the lower basal expression. The concentration, C, of molecule m changes in space and time according to a diffusion degradation equation. For cells arranged in a rectangular lattice this is approximated by:

$$\partial C_i/\partial t = D \triangle C_i\ - \gamma C_i\ +\ \eta(x_S^i)\ /V_c \tag{3. 3}$$

for each voxel i of the lattice containing cell i. D is the diffusion coefficient and γ is the constant degradation rate. Assuming that diffusion is much faster than gene regulation, we use the steady state of the diffusion equation above:

$$0 = D \triangle \hat{C}_i\ - \gamma \hat{C}_i\ +\ \eta(x_S^i)\ /V_c \tag{3. 4}$$

and use a numerical solver for calculating $\hat{C}_i$ in simulations. An important component of the steady state solution is the effective interaction distance, λ, where $\lambda = \sqrt{D/\gamma}\ /\ R$ and R is the radius of the cell [198,199].

When sensing molecule m, cell i checks the local concentration of the signal, i.e., the concentration at its grid point. If the local concentration of m is above a threshold value, θ , then the receptor node is activated, otherwise it is deactivated. Formally, the state of the receptor node x_R^i follows the equation:

$$x_R^i(t+1) = 1, if\ \hat{C}_i(x_S^1(t), \ldots, x_S^N(t)) > \theta, \tag{3. 5}$$

$$x_R^i(t+1) = 0, \text{ otherwise,} \quad (3.6)$$

where $\hat{C}_i(x_S^1(t), \ldots, x_S^N(t))$ is the concentration of m at the position of cell i. We will vary λ and θ to change communication within the tissue. The other model parameters are kept constant for all simulations, with no loss of generality.

Note that while our model assumes diffusion-based cell-cell communication, the effective interaction distance can be shortened such that the system behaves as if signaling were contact-mediated, the latter effectively being a special case of the former. Similarly, though we describe the case of a single signaling molecule, the model could easily be extended to multiple signaling molecules.

3.2.1.4 Simulation Framework

Simulations of our model were implemented in *Biocellion* [200], a high performance computing platform designed for simulation of multicellular systems. At every time step t of the simulation the concentration of signalling molecule m is updated by numerically solving the diffusion equation in section 3.2.1.3, after which the Boolean states of the cells are updated using the computed concentrations.

3.2.2 *Experimental Design*

Using our mathematical model, we performed two sets of experiments. The first set of experiments probed the activity of a cell's receptor node in response to different signaling states of itself and all other cells. The second set of experiments probed the steady state behavior of cells and tissues. Each is described in more detail below.

3.2.2.1 Asocial to Social Transition

In this experiment, we constructed a 3D tissue of N= 4096 (16×16×16) cells, in which the cellular network consists of only the two signaling nodes (Figure 3.3A). The update functions are:

$$x_S^i(t+1) = x_S^i(t) \tag{3. 7}$$

for the signal nodes, and

$$x_R^i(t+1) = 1, if\ \hat{C}_i(x_S^1(t), \ldots, x_S^N(t)) > \theta, \tag{3. 8}$$

$$x_R^i(t+1) = 0, \text{otherwise.} \tag{3. 9}$$

for the receptor nodes. We sampled the effective interaction distance, λ, and the receptor activation threshold, θ, using a fine grid of 271×141 points in the interval [0.5, 6.0] and [1.0, 15.0], respectively, and a coarse grid of 66×200 points in the interval [1.0,14.0] and [1.0,200.0], respectively. Each combination was used to generate a tissue sample. Each tissue sample was initialized with each of a set of four initial conditions and updated for a single time step. The four initial conditions demonstrate the cases in which a single cell i has either $x_S^i(t=0) = 0$ or $x_S^i(t=0) = 1$ and all other neighboring cells $j \neq i$ have either $x_S^j(t=0) = 0$ or $x_S^j(t=0) = 1$ (Figure 3.3B). For all four initial conditions, $x_R^i(t=0) = x_R^j(t=0) = 0$. After the single update of the Boolean functions, the value of $x_R^i(t=1)$ is recorded.

3.2.2.2 Cellular and Tissue behavior

In this experiment, we constructed multiple tissues of N= 4096 (16×16×16) cells, in which the cellular networks are RBNs with n = 12 nodes (10 regulatory nodes and 1 pair of signaling nodes). For each tissue, the cellular Boolean network is instantiated as described in section 3.2.1.1, using in-degree $k = 3$ and bias $p = 0.2113$. These values were chosen to give dynamically critical

cells, the proposed dynamics of living cells [76,152]. It should be noted that artificial cells that do not communicate are likely to be critical at these parameter values; however, when communication is introduced, the dynamical regime of the cells may be altered, meaning that it may no longer be critical.

For each tissue, random initial states are chosen for all $N \times n$ nodes. Thus, not all cells have the same initial state vector. Using this initial condition, 108 different tissue samples are simulated for 2000 iterations each. These simulations cover a combination of 9 λ and 12 θ values. The specific values for these parameters were chosen such that the different communication regions in the λ, θ parameter space are represented (Figure S3.1). The 9 λ values are incremented by cell radius lengths and the 12 θ values are sampled such that there is 1 in region A1, 1 in S1, 5 uniformly in S2 or A2, 1 in S3, and 1 in A3.

For analysis, we used data from 100 tissues that are dynamically stable and have reached an attractor in all tissue samples within 2000 iterations. A tissue sample reaches an attractor once all cells collectively revisit the same cell states, i.e., the set of all nodes in the tissue revisits the same states. Formally, this occurs when there exists a $\tau > 0$ such that:

$$X_i(t) = X_i(t-\tau), i = 1, \ldots, N. \quad (3.\ 10)$$

The tissue attractor is defined as the vector of all cell states that occurred between time $t - \tau$ and t:

$$A^{T_{\lambda,\theta}} = [A^{C_1}, A^{C_2}, \ldots, A^{C_N}], \quad (3.\ 11)$$

where

$$A^{C_i} = \{X_i(t-\tau), X_i(t-\tau+1), \ldots, X_i(t-1)\}, i = 1, \ldots, N. \quad (3.\ 12)$$

We refer to the stable dynamics of a cell, A^C, as the cellular attractor (CA) (see Figure 3.2 for an example in non-communicating cells). Note that typically, attractors are fully ordered sets under circular permutation; however, we do not consider state order. In practice, we have observed the state order is never the sole discerning feature between two cellular attractors. There exists a set of unique CAs, $\mathcal{A}_1, \mathcal{A}_{2,\dots,}\mathcal{A}_M$ for every tissue such that for every $i \in \{1,\dots,N\} \exists\ j \in \{1,\dots,M\}$ such that $A^{Ci} = \mathcal{A}_j$. We will use the cellular attractor to characterize cellular behavior in our analyses.

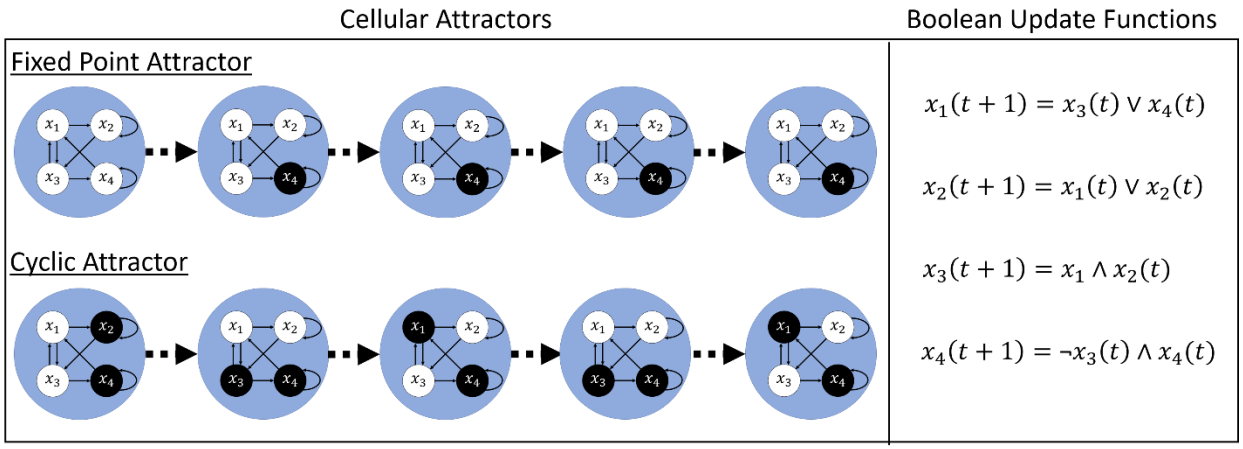


Figure 3.2. Examples of cellular attractors for a non-communicating cell. Two possible cellular attractors (left panel) for the given Boolean network (right panel) are shown. One cellular attractor (top) is a fixed point attractor in which the network stays in the same state after entering it. The other cellular attractor (bottom) is a cyclic attractor consisting of two states between which the network alternates.

3.3 Results and Discussion

3.3.1 *Asocial to Social Transition: When does signaling become communication?*

Before asking how cellular and tissue behavior are affected by changing communication within the tissue, we first had to understand how communication is changed by the effective interaction distance, λ, and receptor activation threshold, θ. Even though signals may be secreted by a cell's neighbors, the signal may not travel far enough or be concentrated enough to activate the neighboring cell's receptor. Additionally, a cell's own signal may mask that of its neighbors [103], effectively negating intercellular communication. Therefore, we investigated which λ and θ are necessary for cells to be able to influence each other. To explore this, we ran simulations to check receptor response to different combinations of self and neighbor signaling states under different λ, θ values. The simulations are described in detail in the section 3.2.2.1. Briefly, the cells in a tissue are composed of a pair of Boolean signaling nodes, a receptor, x_R, and a signal node, x_S (Figure 3.3A). For different λ, θ values, we recorded the immediate response of a single cell's (cell i) receptor, x_S^i, to four different initial conditions of cell i's and all other cells' j signal nodes, x_S^i and x_S^j, $j \neq i$ (Figure 3.3B). The initial conditions are:

(1) All cells' signal nodes are OFF: $x_S^i = x_S^j = 0$

(2) Cell i's signal node is OFF and all other cells' signal nodes are ON: $x_S^i = 0$, $x_S^j = 1$

(3) Cell i's signal node is ON and all other cells' signal nodes are OFF: $x_S^i = 1$, $x_S^j = 0$

(4) All cells' signal nodes are ON: $x_S^i = x_S^j = 1$

By examining whether cell i's receptor node turns ON in each of these scenarios, we are determining its level of social interaction. In fact, we find six distinct signaling regions across our λ, θ parameter space, which can be split into social and asocial categories (Figure 3.4). Because

of the interplay of λ and θ in receptor activation, the boundaries between these six regions are curves in the λ, θ space. That is, as the effective interaction distance becomes longer, more signal will reach a cell and its receptor can have a higher receptor activation threshold and still be activated.

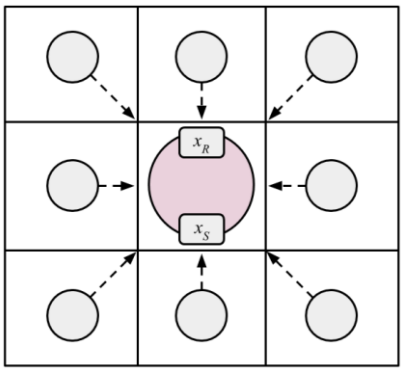
Asocial behavior occurs in regions A1, A2, and A3 of the parameter space (Figure 3.4). In region A1, cell i's receptor turns ON regardless of whether itself or other cells are generating signal. This occurs when θ is so low that the basal secretion from cells in cell i's neighborhood activates its receptor. As one might expect, as the receptor activation threshold increases, the effective interaction distance must increase in order for a cell to receive enough signal to have this always ON behavior. We will refer to this region as the ON region. In contrast, in region A3, cell i's receptor remains OFF regardless of whether itself or its neighbors are generating signal. Here, θ is so high that even the maximum amount of signal cannot activate the receptor. We will refer to this region as the OFF region. The last asocial region, region A2, can be described as having self-talk behavior. Cell i's receptor is only activated by its own signal, i.e. $x_R^i(t+1) = x_S^i(t)$. We will refer to this region as the SELF region. Thus, in regions A1, A2, and A3, cells are behaving completely independently of each other.

Signaling becomes communication and cells are affected by their social environment in regions S1, S2, and S3 of the λ, θ space (Figure 3.4). In the extreme case, in region S2, the state of cell i's receptor is solely determined by the signal secreted by its neighbors. In regions S1 and S3, cell i's receptor is affected by both self and neighbor signal. Specifically, in region S1 cell i's receptor can be activated by *either* itself or neighbors; in region S3, cell i *must* receive signal from both itself and neighbors in order to activate its receptor. For each of these regions, more neighbor signal is required as θ is increased.

While these results were obtained from experimental simulations, one can also analytically derive the boundary curves for each social region using an approximation of the steady state equation for molecule concentration (see section 3.5). This analysis also reveals that the relative positioning of the boundary curves is independent of many system parameters, including: the basal secretion rate ($\eta(0)$), the active secretion rate (α), the diffusion coefficient (D), the radius of cells (R), the spacing between cells (H), and the size of the population (N). So long as these parameters are constant within a population, the relative distribution of the regions and their boundaries will not change.

Here, we have only considered the extreme case in which every neighbor cell is either signaling or not. However, the switch in social behavior occurs exactly at these extremes. Within each region, the particular response is dependent on both the fraction of signaling neighbors and their spatial configuration, which was explored by Maire et al. [198] in a similar theoretical framework. In examining the role of the secretion rate and receptor activation threshold in determining the relative autonomy of bacteria, they demonstrate the same social dynamics as we have.

(A)



(B)

Condition	$x^i_S(t = 1)$	$x^j_S(t = 1)$
1	0	0
2	0	1
3	1	0
4	1	1

Figure 3.3. Experimental design for testing the asocial to social transition of cells. **(A)** Population of simple cells with only 2 nodes, x_S and x_R. **(B)** Initial conditions for the simple 2-node experiment. $x_S^i(t = 1)$ is the initial state of cell i's signaling node and $x_S^j(t = 1)$ is the initial state of all other cells' j signaling nodes.

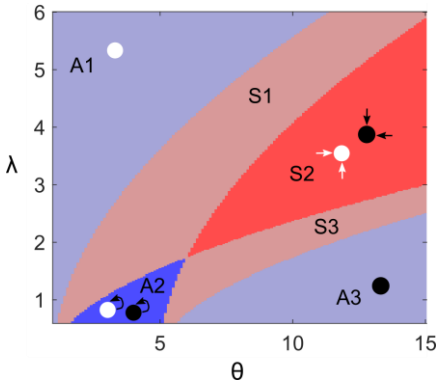


Figure 3.4. The six distinct communication regions in the λ, θ space. Each region corresponds to a distinct signaling behavior of any cell i in a tissue sample with a given λ, θ value. In the asocial regions (blue), cell i's receptor is always ON (A1 - light blue), responds to self-signal (A2 - dark blue), or always OFF (A3 - light blue). In the social regions (red), cell i's receptor can be activated by self or neighbor signal (S1 - light red), only neighbor signal (S2 - dark red), or both self and neighbor signal (S3 - light red). For easier visualization, only $\lambda \leq 6$ is shown. See Figure S3.1 for the wider parameter range used in simulations

3.3.2 *The impact of communication on cellular and tissue behavior*

We then investigated how cellular and tissue behavior can change between and within the six signaling regions. For this, we simulated tissues with different λ, θ values and analyzed the resulting cellular and tissue dynamics. The details of these simulations are described in section 3.2.2.2. Briefly, we generated 100 unique tissues, each with a particular cellular Boolean network. The cellular BNs consist of 12 nodes (10 regulatory, 1 signaling pair), allowing for a variety of possible cellular dynamics. We simulated each tissue with 108 λ, θ parameter values (12 $\lambda \times 9\ \theta$ values) that were sampled from across the different communication regions (Figure S3.1). The initial conditions of a tissue sample are the same for all λ, θ pairs, so that any variations in dynamics is attributable to changes in communication only. For analysis, we primarily utilize cellular attractors, which we consider to be distinct cellular behaviors.

3.3.2.1 Does changing communication cause cells to alter their behavior?

It is clear that signaling between cells can influence cellular behavior. Here we investigate how the magnitude of that influence varies with changes in communication. Specifically, we tested whether the likelihood of a significant change to a cell's behavior is dependent on the degree of communication. For a given tissue sample, we identified which cellular attractor each of its 4096 cells occupied. Representing the tissue by its tissue attractor for each λ, θ pair, we have:

$$A^{T_{\lambda,\theta}} = [A^{C_1}, A^{C_2}, \ldots, A^{C_N}] \tag{3.11}$$

where A^{C_i} is the cellular attractor occupied by cell i (see section 3.2.2.2). To quantify the effect of changing communication on cellular behavior, we measured the fraction of cells in a tissue sample that are in a different cellular attractor compared to when no communication occurs. Thus, we use the three asocial regions (ON, OFF, and SELF) as references for changes in cellular behavior in

tissues in the social regions (S1,S2,S3). To compare a social tissue sample to any one asocial reference, we use the normalized Hamming distance:

$$D_H(A^{T_{\lambda,\theta}}, A^{T_{\lambda\prime,\theta\prime}}) = \frac{\sum_{i=1}^{N} \chi(A^{T_{\lambda,\theta}}(i) = A^{T_{\lambda\prime,\theta\prime}}(i))}{N} \quad (3.\ 13)$$

where $A^{T_{\lambda\prime,\theta\prime}}$ is one of $A^{T_{\lambda^{ON},\theta^{ON}}}$, $A^{T_{\lambda^{OFF},\theta^{OFF}}}$, or $A^{T_{\lambda^{SELF},\theta^{SELF}}}$- the CA sets corresponding to the three asocial regions. $\chi[a]$is an indicator function that outputs a one if a is true and zero if a is false. We use the minimum of these three values:

$$b_{\lambda,\theta} = \min(D_H(A^{T_{\lambda,\theta}}, A^{T_{\lambda^{ON},\theta^{ON}}}), D_H(A^{T_{\lambda,\theta}}, A^{T_{\lambda^{OFF},\theta^{OFF}}}), D_H(A^{T_{\lambda,\theta}}, A^{T_{\lambda^{SELF},\theta^{SELF}}})) \quad (3.\ 14)$$

as our metric of cellular change so that we do not conflate shifts away from the asocial references with shifts between the references. By taking the minimum, $b_{\lambda,\theta}$ effectively measures the amount of change in cellular behavior that is guaranteed to be away from the asocial references.

We find that in 36% of the tissues, cells do not change their CA regardless of how communication is tuned (Figure 3.5A). In all but one of these tissues, a cell's behavior in the social regions is identical to that in the SELF region. Moreover, the SELF region in these tissues is identical to that in either the ON or OFF regions, indicating that the signal molecule is either always secreted or never secreted. In the remaining tissue, there are distinct ON, OFF, and SELF regions. A cell's behavior in the social regions is identical to that in the ON region except for in region S3 where it is identical to that in the SELF region. In all these instances, cells maintain asocial behavior even when communication occurs.

In the other 64% tissues, some fraction of cells entered a CA in the social regions that is different than that of the asocial reference. There are three general trends in $b_{\lambda,\theta}$ for a tissue. Within the social regions, $b_{\lambda,\theta}$can either be equal across all λ,θ values (Figure 3.5B), have a single peak that roughly follows a curve in λ, θ space (Figure 3.5C), or two peaks that roughly follow curves

in λ, θ space (Figure 3.5D). Additionally, these tissues can behave identically to the ON, OFF, and SELF regions even when the effective interaction distance and receptor activation threshold place the tissue in the social regions (Figure S3.2A). For each of these general trends, the particular response of cells is variable between tissues (Figure S3.3). Interestingly, the greatest increase in $b_{\lambda,\theta}$ occurs within a short neighborhood of communication (λ= 4,5) (Figure 3.6). Overall, it is clear that changing communication between cells can lead to measurable changes in a cell's behavior, with significant diversity in the responses of individual tissues.

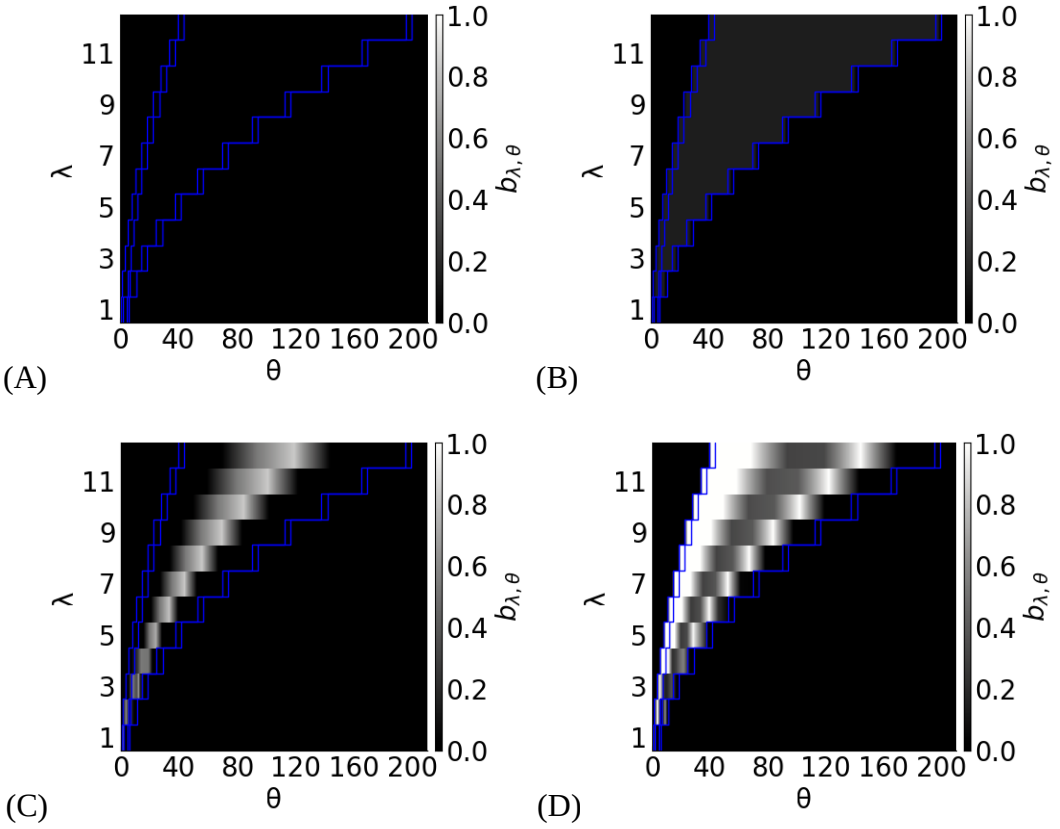


Figure 3.5. Examples of different trends in how $b_{\lambda,\theta}$ changes with the degree of communication. Each subplot represents data for a single tissue at sampled λ, θ. The color indicates the value of $b_{\lambda,\theta}$, with interpolation of values between sampled θ for each λ. Blue lines represent the boundaries of communication regions as measured in section 3.3.1.

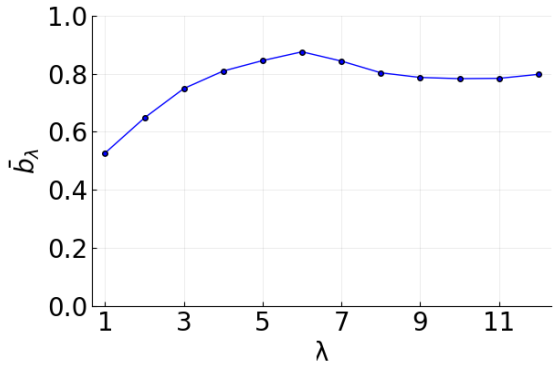


Figure 3.6. Maximum $b_{\lambda,\theta}$ as a function of λ. For each λ value, the maximum $b_{\lambda,\theta}$ value is calculated, b_λ. The blue line shows the mean of the maxima, $\bar{b}_\lambda$, across all 100 tissues for each λ.

3.3.2.2 Do cells behave differently when they communicate?

Next, we considered whether cell-cell communication can alter the range of possible cellular behaviors. In other words, is cellular behavior limited or expanded by communication? Counting the number of all cellular attractors observed for a given cellular BN in either asocial cells (regions A1,A2,A3) or social cells (regions S1, S2, S3), we found social cells can exhibit different numbers of CAs than asocial ones (Figure 3.7A). In both social or asocial cells, roughly 50% of cellular BNs are capable of being in ~5 or fewer CAs; however, the tail of the distribution for social cells is heavier and longer than that for asocial cells (Figure 3.7B).

Given the discrepancy between the number of cellular attractors available to cells in social versus in asocial tissues, the set of CAs in each region must be distinct, i.e., cells that communicate behave differently than cells that do not. By comparing CAs between the asocial and social regions, we found that cellular BNs can both lose or gain CAs when communicating (Figure 3.8A,B). Losing or gaining CAs is not rare, with 66% cellular BNs missing at least one CA (Figure 3.8A)

and 34% cellular BNs gaining at least one CA (Figure 3.8B), with some cellular BNs both missing and gaining CAs. To check that the missing or novel CAs are not infrequent cellular behaviors, and therefore less likely to be biologically meaningful, we measured their frequencies within a given tissue sample, reporting the maximum across all communication parameter values (Figure 3.8C,D). Missing CAs can be very frequent in asocial tissues and novel CAs can range from infrequent to very frequent in social tissues.

While missing CAs are missing from the entire social communication space, the expression of novel CAs is not evenly distributed across interaction distances and receptor threshold values. Some novel CAs are observed in the entire space but others are only observed at particular λ,θ values. Moreover, we found that, on average, a cellular BN expresses the largest fraction of its novel CAs at λ= 3 (Figure 3.9A). Additionally, the maximum fraction of cells in a novel CA within a tissue sample increases with λ until λ= 5 (Figure 3.9B). Thus, a cell will have the highest chance of being in the most novel CAs when communication is limited to the distance one or two neighboring cells.

In order to understand how novel CAs can arise in social cells, we considered how they are related to other CAs. Per the definition of a cellular attractor, a CA consists of a set of cellular states. In the case of asocial cells, when cells are essentially closed systems, a cell state X will only occur once and will always have the same following cell state X'. Therefore, the CAs of cells within a particular asocial region are composed of completely disjoint sets of states. When communication is introduced, cells become open systems and the composition of their CAs is no longer constrained. Thus, it is possible that the novel CAs are variations of existing CAs. To get an idea of how CAs that arise in social cells are related to those in asocial cells, we used the set intersection between CAs to calculate the overlap between each novel CA and each asocial CA.

Using this method, novel CAs can be classified into three broad categories: modified, combined, and true novel. A modified CA is one that has a non-empty set intersection with one (and only one) asocial CA with the addition of other cell states. These additional cell states are transient states in asocial cells that become stabilized in social cells. A combined CA is one that has a non-empty set intersection with more than one asocial CA with or without the addition of cell states. The final category, true novel, describes a CA that has an empty set intersection with all asocial CAs. These true novel CAs are composed entirely of states that were unstable transients in asocial cells. Notably, a cellular BN can have one, two, or all three types of novel CAs. Overall, we found that combined CAs are the most common category, being present in the highest fraction of tissues (Figure 3.10A) and usually the highest fraction of the novel CAs for any of one tissue (Figure 3.10B). Modified CAs are the next most common followed by true novel.

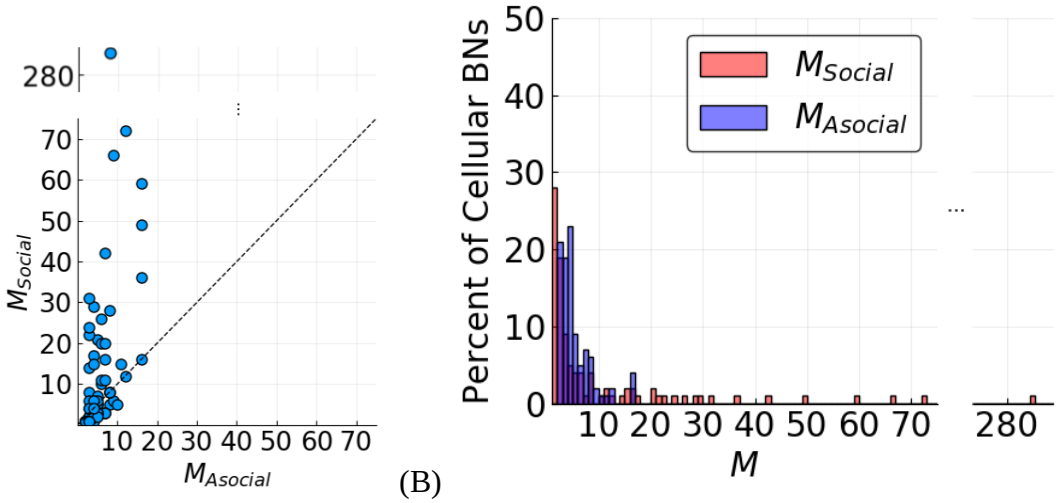


Figure 3.7. Numbers of cellular attractors observed in asocial (A1,A2,A3) and social (S1,S2,S3) regions of the λ,θ parameter space. $M_{Asocial}$ and M_{Social} are the number of unique CAs observed in the asocial and social λ,θ regions, respectively. **(A)** $M_{Asocial}$vs. M_{Social}. **(B)** Histograms of $M_{Asocial}$ and M_{Social}.

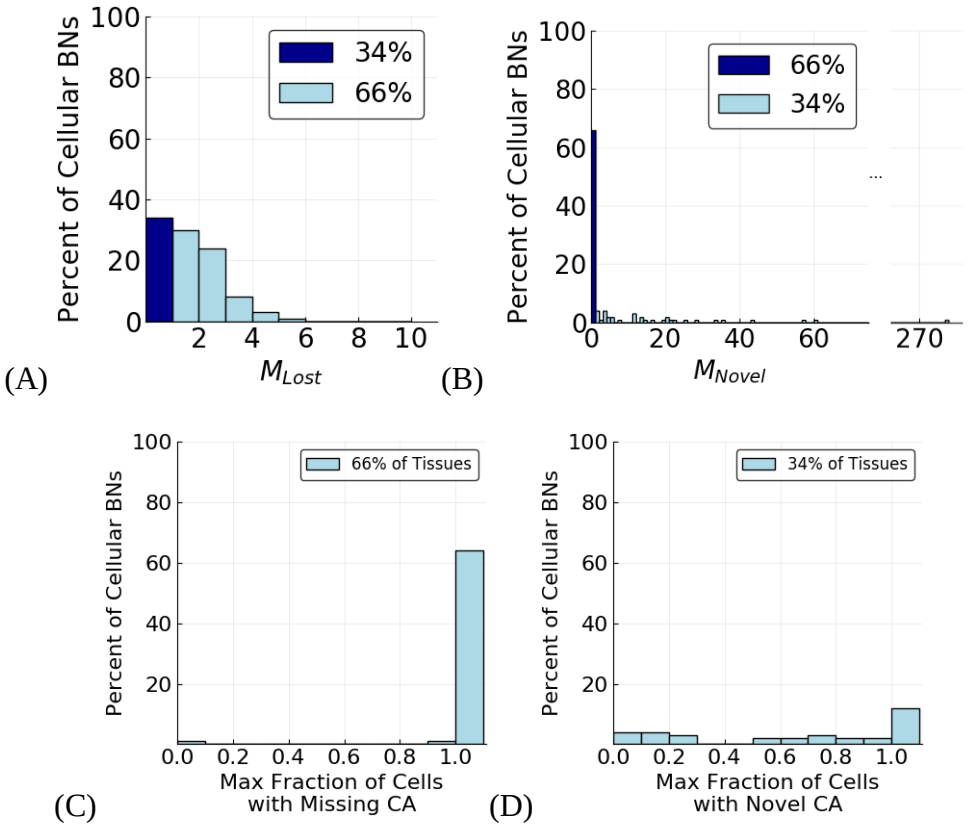


Figure 3.8. Likelihood of observing Lost or Novel CAs in tissues. **(A)** Distribution of the number of unique Lost CAs in a single tissue. 66% of cellular BNs are missing at least one CA (light blue) and 34% are not missing any CAs (dark blue). **(B)** Distribution of the number of unique Novel CAs in a single tissue. 34% of cellular BNs gain at least one CA (light blue) and 66% do not gain any CAs (dark blue). **(C)** Distribution of the maximum frequency of Lost CAs in an asocial tissue sample. **(D)** Distribution of the maximum frequency of Novel CAs in a social tissue sample.

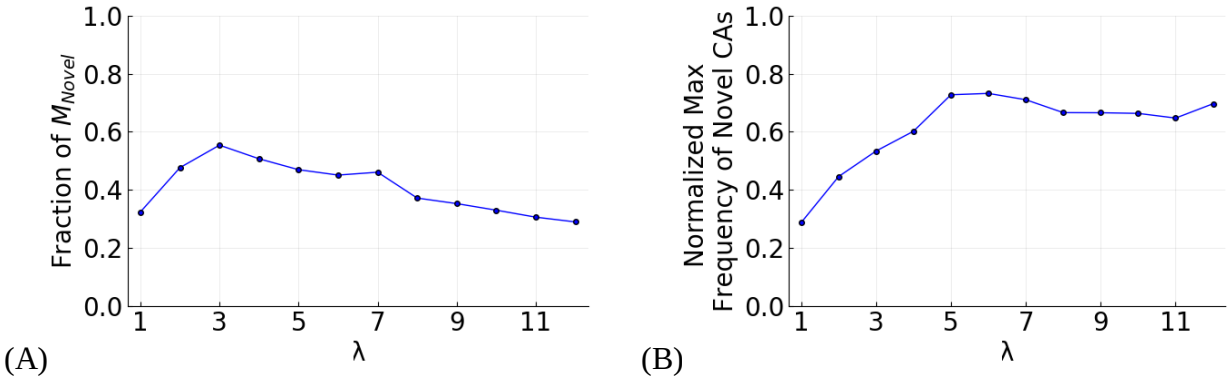


Figure 3.9. Relationship between λ and the expression of novel CAs. **(A)** Availability of novel CAs as a function of λ. M_{Novel} is the number of unique Novel CAs for a given tissue. The mean of the fraction of M_{Novel} Novel CAs expressed at each λ is plotted. **(B)** Frequency of Novel CAs as a function of λ. The mean of the normalized maximum fraction of cells in novel CAs within a tissue sample for a given λ is plotted.

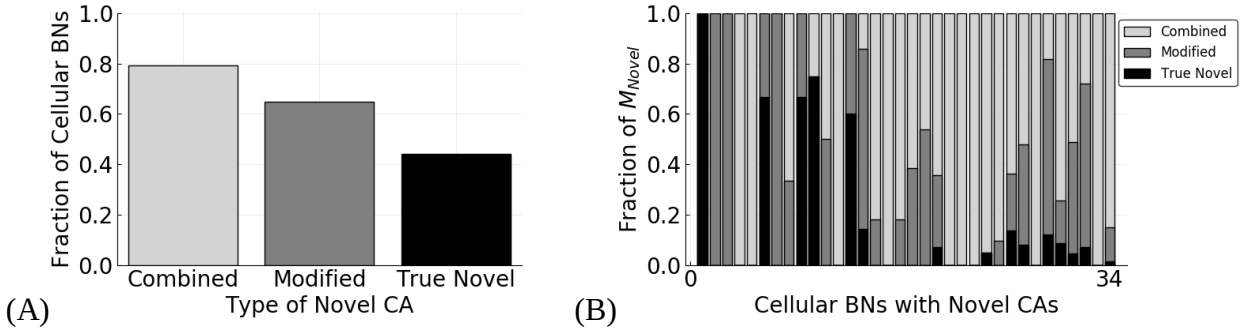


Figure 3.10. Characterization of Novel CAs. **(A)** Fraction of cellular BNs with each of the three categories of Novel CAs. **(B)** Distribution of each category of Novel CA within each tissue.

3.3.2.3 Does the composition of a tissue change as communication is added?

Here we turn our focus to the effects of changing communication on population behavior. We began by testing whether a tissue has any meaningful change to its behavior in response to changes in communication. To quantify the effect of changing communication, we measure changes in which CAs are expressed by cells as we did when testing for the impact on cellular behavior.

However, rather than tracking individual cells, we look at the macroscopic property of CA distributions within the tissue. Since lower-level changes to cellular identity do not guarantee higher-level changes, changes in individual cells may not be reflected in changes to the tissue.

To construct a CA distribution for a given tissue sample, we calculate the fraction of cells in the tissue with each CA, which gives:

$$D_{\lambda,\theta} = \{d_{\lambda,\theta}^{1}.\ d_{\lambda,\theta}^{2}, \ldots, d_{\lambda,\theta}^{M}\}, \tag{3. 15}$$

where

$$d_{\lambda,\theta}^{i} = \frac{\sum_{j=1}^{N} \chi(CA_j = i)}{N} \tag{3. 16}$$

and M is the number of unique CAs for that tissue across all interaction distances, λ, and receptor thresholds, θ. To measure changes to a CA distribution, we compare distributions between tissue samples using the symmetric Kullback-Leibler (KL) divergence. The symmetric KL divergence for P and Q is given by:

$$D_{KL}(P, Q) = D_{KL}(P||Q) + D_{KL}(Q||P) \tag{3. 17}$$

which is the sum of the two KL divergences:

$$D_{KL}(P||Q) = -\sum_{x \in X} P(x) \log\left(\frac{Q(x)}{P(x)}\right) \tag{3. 18}$$

and similarly defined, $D_{KL}(Q||P)$. Here we use the natural logarithm, though any base may be used for the logarithm. If P and Q are identical distributions, then the symmetric KL divergence will be 0 and will increase as P and Q diverge.

In our calculations, P is the CA distribution of a tissue, $D_{\lambda,\theta}$, and Q is the CA distribution of tissues in one of the three asocial regions: $D_{\lambda^{ON},\theta^{ON}}$, $D_{\lambda^{OFF},\theta^{OFF}}$, $D_{\lambda^{SELF},\theta^{SELF}}$. As in section

3.3.2.1, we are using the asocial regions as references to quantify the effect of changing communication, taking the minimum KL divergence to guarantee the maximal divergence from any reference. Therefore, we define:

$$\widehat{D}_{KL}(\lambda,\theta) = \min(D_{KL}(D_{\lambda,\theta,}D_{\lambda^{ON},\theta^{ON}}), D_{KL}(D_{\lambda,\theta,}D_{\lambda^{OFF},\theta^{OFF}}), D_{KL}(D_{\lambda,\theta,}D_{\lambda^{SELF},\theta^{SELF}})) \quad (3.19)$$

which we used as our metric of change in tissue composition.

As expected, the 36% of tissues that showed no change in cellular behavior also show no change to tissue composition (Figure 3.11A). Even though communication occurs within these tissues, they behave the same as asocial tissues. For the other 64% of tissues, there is a mixed relationship between the likelihood of cellular change and changes in tissue composition. In some tissues, there is a clear positive correspondence (Figure S3.4A). In others, there seems to be little change in tissue composition regardless of switching cellular behavior (Figure S3.4B). Most interestingly, some tissues have little change in tissue composition until some percentage of cells switch behavior (Figure S3.4C).

As with the behavior of individual cells, tissue composition is clearly a function of the degree of communication (Figure 3.11B-D). The broad regions of asocial and social behavior are reflected in the shifts of tissue composition. Even so, the response of each tissue is unique, with different responses in how tissue composition changes within the social regions of λ and θ. The same three general trends observed for changes in cellular behavior (flat, single peak, double peak) are also observed for changes in tissue composition. In many cases, the tissue composition of the ON, OFF, or SELF regions creeps into the social regions, minimizing the communication parameter space where distinct tissue compositions can occur (Figure S3.2B). Looking across all of the tissues, we find that, on average, tissue composition changes most rapidly as the effective

interaction distance increases beyond a single cell, peaking by λ= 5 and then leveling out (Figure 3.12).

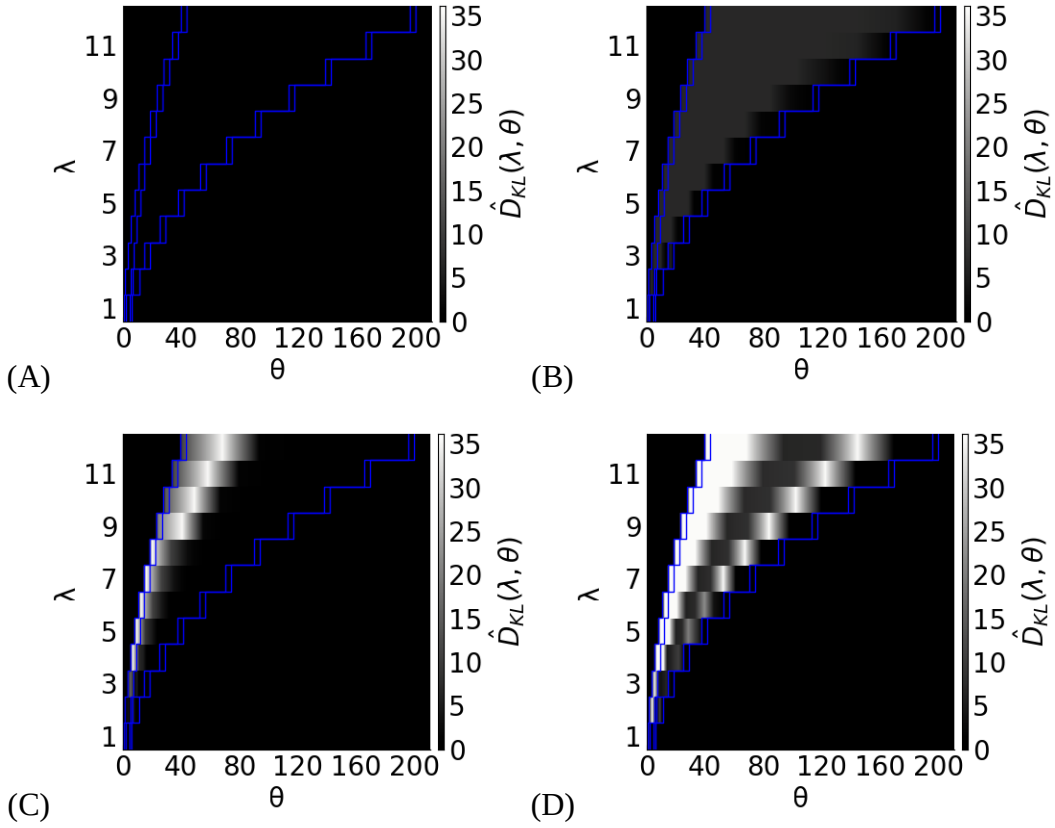


Figure 3.11. Examples of how different tissues change CA composition with the degree of communication. $\widehat{D}_{KL}$ increases with increasing change in composition. Each subplot shows the $\widehat{D}_{KL}$ for one tissue across sampled λ, θ values. **(A)** A tissue with no change in composition. **(B)** A tissue with a flat $\widehat{D}_{KL}$ response to changing communication. **(C)** A tissue with a single peak in its $\widehat{D}_{KL}$ response. **(D)** A tissue with two peaks in its $\widehat{D}_{KL}$ response. Blue lines represent the boundaries of communication regions as measured in section 3.3.1.

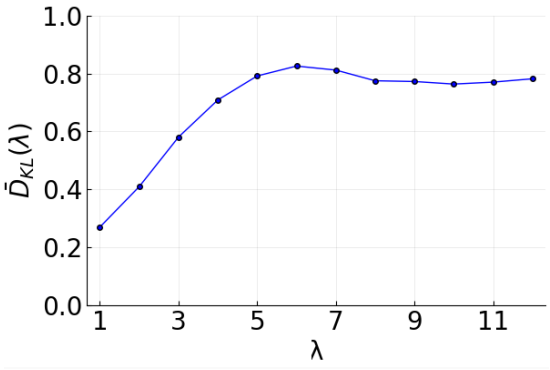


Figure 3.12. $\widehat{D}_{KL}$ as a function of λ. For a given λ value, the maximum $\widehat{D}_{KL}$ was calculated for each tissue and then averaged across tissues, giving $\overline{D}_{KL}(\lambda)$.

3.3.2.4 Does the diversity of a tissue depend on communication?

Taking a deeper examination of how tissue composition is affected by communication, we analyzed the diversity within tissues. Population diversity is a functionally important property of cellular populations, and is tightly regulated, whether for generating homogeneity or heterogeneity [158,164,176]. Using the CA distributions of tissues, we calculated the entropy of the distribution as a measure of diversity - completely homogenous tissues having the lowest entropy and tissues with an equal number of every CA having the highest entropy. Tissue diversity is thus defined as:

$$H_{\lambda,\theta} = -\sum_{i=1}^{M} d_{\lambda,\theta}^{i} \log(d_{\lambda,\theta}^{i}) \ . \qquad (3.20)$$

First, we examined the specific case of homogeneous tissues, in which $H_{\lambda,\theta}$= 0. We found a clear trend between the fraction of tissue samples that are homogenous and the degree of communication (Figure 3.13A). At λ= 2 there is an immediate drop in the number of tissues with only one cell type. Remarkably, it is also the global minimum of the fraction of homogeneous tissues. As the effective interaction distance and receptor activation threshold are increased along

the curve, the fraction of homogeneous tissues increases. Notably, this fraction is both highest and lowest within the social regions, but peaks or dips at a low λ value (Figure 3.13B).

Looking more generally at all values of entropy, we found that communication has a strong effect on tissue diversity. Of course, for those tissues in which tissue composition does not change, $H_{\lambda,\theta}$ is identical to that of one of the asocial regions. In the other tissues, entropy varies within the social region along λ, θ curves, taking values that are higher, lower, and/or the same as in asocial regions (Figure 3.14A-D). Similar to how changes in cellular behavior do not guarantee changes in tissue composition, changes in tissue composition do not guarantee changes in tissue entropy. In fact, even when tissue composition differs from asocial tissues, it is possible for entropy within the tissue to largely remain the same (Figure S3.5A). We also observed tissues in which changes in tissue composition and entropy are uncorrelated, negatively correlated, or positively correlated (Figure S3.5B-D).

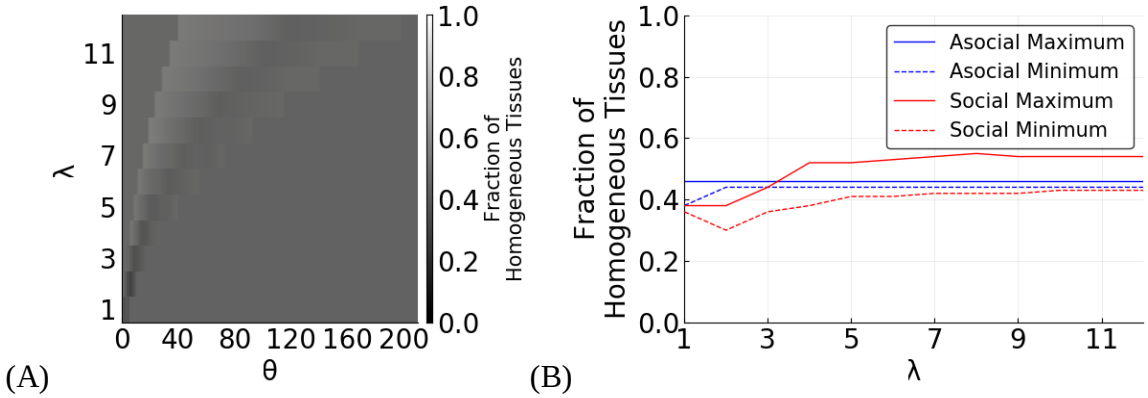


Figure 3.13. Fraction of tissue samples that are homogeneous as a function of λ and θ. **(A)** Heatmap of fraction of homogeneous tissues as a function of λ and θ. **(B)** Maximum (solid) and minimum (dashed) fraction of homogeneous tissues at each λ value, split by asocial (blue) and social (red) regions. Both (near) global max and global min occur in social regions at $\lambda \leq 4$.

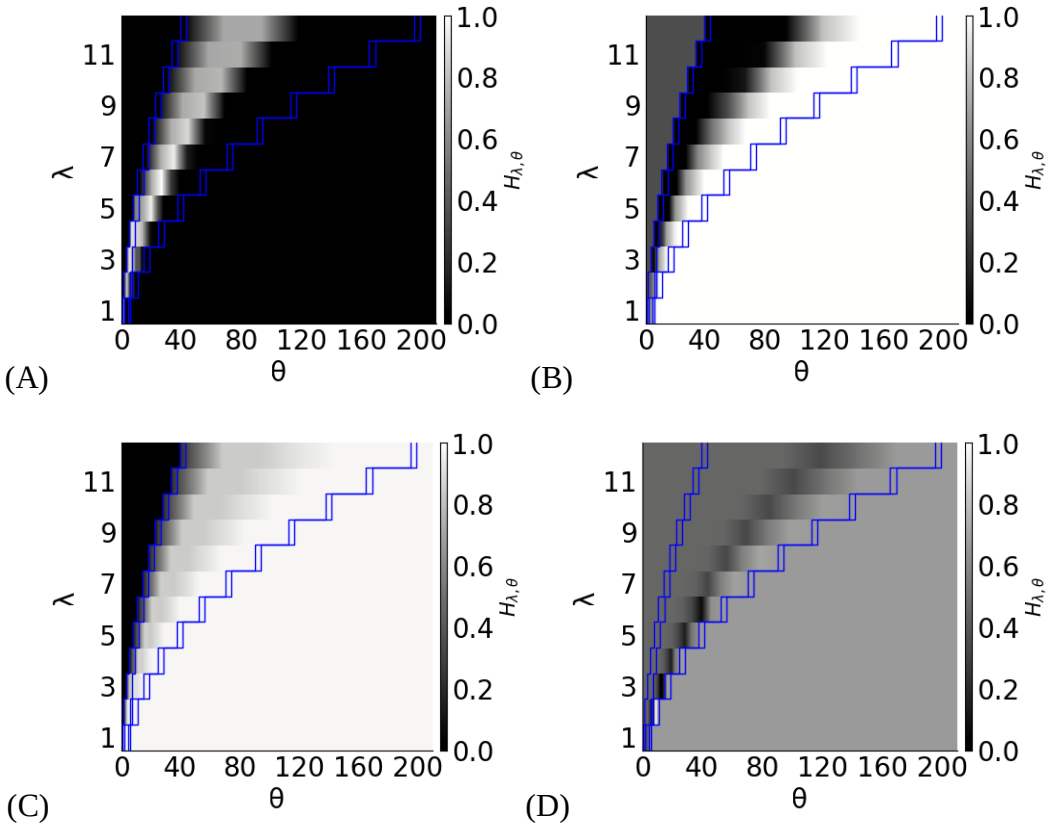


Figure 3.14. Examples of how tissue entropy changes with changing communication. Each subplot represents the entropy, $H_{\lambda,\theta}$, of one tissue at different λ,θ values. Tissues can have entropy that is higher **(A)**, lower **(B)**, or in-between **(C)** that of tissues in the asocial regions. Some tissues can have entropy that is higher, lower, and in-between that of the asocial regions **(D)**.

3.4 DISCUSSION

Communication is a key component of living systems, providing a means by which cells can coordinate their activities to achieve collective goals and accomplish complex functions. Understanding the range and magnitude of the effects of communication within populations advances our understanding of how communication systems have evolved and how they function

in extant organisms. It also advances our ability to modify, manipulate, and repair existing or synthetic systems.

In this work, we have used an *in silico* model of 3-dimensional cellular populations to explore the effects communication can have on cellular and population behavior. By modeling a generalized system, we aimed to explore the effects of communication without the bias of systems evolved for specific functions. We focused on two signaling parameters, the effective interaction distance between cells, λ, and the receptor activation threshold, θ. These parameters have previously been shown to affect the degree of communication between cells [103,198]. In addition, both parameters are variable within and between populations [201-203]. The effective interaction distance between cells can be altered by the biochemical properties of the signaling molecule, binding properties of the signal receptor, and the contents of the extracellular environment, all of which can influence the stability, concentration, and diffusion rate of the signal molecule. The receptor activation threshold can be altered by the biochemical properties of the receptor and the level of receptor expression by the cell, both of which can change the likelihood of receptor activation in the presence of signaling molecule.

Using our model, we have shown that signaling by cells does not guarantee communication. Moreover, depending on λ and θ, cells can operate in very different social regimes. They can be completely asocial, either with receptors always ON/OFF or solely dependent on self-generated signal. Alternatively, they can be social, either with receptor activation only dependent on neighbor-generated signal or on some combination of self- and neighbor-generated signal. Similar social regimes have been shown by Youk et al. [198] who explored the role of the secretion rate and receptor activation threshold in a similar framework.

Thus, changing λ and θ are possible mechanisms by which populations that use the same signaling framework can exhibit different social behaviors.

When cells do communicate, they coordinate by regulating cellular behavior within the population. When a cell receives a signal, it uses the signal as a cue to change its behavior. We found that the fraction of cells within a population that change behavior due to communication is highly variable with λ and θ. Furthermore, we found that the set of behaviors that cells can exhibit is not consistent between asocial and social cells. Cellular behaviors can be lost or gained due to cell-cell communication. Interestingly, Damiani et al. [105] showed that new behaviors can also arise in communicating populations when changing coupling strength, i.e. the number of unique signals used in communication. Fundamentally, these findings mean that asocial cells and socials cells are operating within distinct sets of functionalities. That is, cells can not only use cell-cell communication to coordinate the expression of different cellular behaviors, but whether a cellular behavior is expressed at all. One repercussion of this is that disrupting communication in social populations can either result in the removal of social behaviors or the appearance of cellular behaviors observed in asocial systems. Similarly, introducing communication to an asocial population can result in the loss of behaviors or the appearance of unpredicted behaviors. Notably, since we are working with randomly generated cellular networks, our results show that the loss and gain of cellular behaviors is an emergent property of populations in which cells communicate rather than a product of evolutionary fine-tuning.

We further examined how distinct cellular behaviors can arise in social cells, finding that they can either be variants of asocial behaviors, combinations of asocial behaviors, or completely distinct from asocial behaviors. Since cellular behaviors are often linked to specific functions, there may be a functional interpretation to these categories. Specifically, social behaviors that are

variations or combinations of asocial behaviors may be functionally similar and social behaviors that are distinct may have distinct functions.

While the behavior of individual cells is regulated by intercellular signals, cell-cell communication is ultimately a population phenomenon. Population-level properties are also regulated by communication as cellular level changes culminate in population level changes. We focused on the distribution and diversity of cellular behaviors within a tissue (tissue composition, tissue diversity) as general population properties that are often important in achieving collective behavior, especially division of labor. We have shown that both tissue composition and diversity can be tuned by changing the effective interaction distance and receptor activation threshold. Interestingly, though changes in the behavior of individual cells is a requirement for changes in tissue composition, the relationship between the two is not the same for all tissues. Similarly, the relationship between tissue composition and tissue diversity is not the same across tissues. In some tissues, diversity will not change with interaction distance and receptor threshold even if tissue composition does. In others, diversity decreases, in many cases becoming homogeneous. In fact, we have shown that when communication is added, the percent of behaviorally homogeneous tissues is lowest at short interaction distances and highest after that. Interestingly, Damiani et al. [105] showed that the percent of homogeneous tissues increases with the coupling strength. These findings together indicate that a stronger connection between cells, whether it be through greater coupling to one cell or more communication partners, can strengthen the synchronization between them. In complete contrast to this, we found tissues in which cell-cell communication leads to an increase in diversity. What determines whether a tissue will homogenize or diversify is unclear, though it is likely related to the structure of the cellular network.

For each of the cellular and tissue properties we have discussed, we have found that not only do different tissues have different responses to changing cell-cell communication, but that a single tissue is capable of distinct responses within the social regions of the λ, θ communication parameter space. However, there is one overarching trend that can be seen across both cellular and population metrics. As the neighborhood of interaction is extended beyond two neighboring cell lengths, there is a diminishing response to cell-cell communication. On average, both the sharpest change and greatest magnitude of response is observed within this range. Effectively, populations of cells do not need to interact beyond a range of one or two cells to utilize the range of effects that cell-cell communication grants.

Overall, there is a clear difference between the cellular and population behavior of communicating and non-communicating populations. Surprisingly, genetically diverse populations all exhibit the same set of general trends in their response to cell-cell communication. However, there is still flexibility in the specific responses of different populations across cellular and population properties. By tuning their signaling parameters, cells can switch between broad types of asocial and social behavior. Even within social populations, there is heterogeneity and flexibility in response to changes in signaling parameters. Furthermore, signaling parameters can vary across space or across signaling molecules. Thus, social behavior can vary across a single population or across functional components of a single cell. Additionally, external factors may modify signaling parameters, allowing for cells to utilize environmental cues to alter their behavior or for an external agent to manipulate the system. This diversity enables specialization for the unique selective pressures of any given population, tuning their signaling parameters over evolutionary time or within the lifetime of a cell.

3.5 APPENDIX

Analysis of Social Boundary Curves. The diffusion equation for the concentration of molecule m at the surface of cell i is given by:

$$\partial C_i/\partial t = D \triangle C_i - \gamma C_i + \eta(x_{i,S}) /V_c \tag{3.3}$$

where C_i is the concentration at cell i, D is the diffusion coefficient, γ is the rate of degradation, $\eta(x_{i,S})$ is the secretion rate of cell i, and V_c is the volume of a cell. If we assume that cells are identical spheres and that boundary conditions are not periodic, then the steady-state concentration, $\hat{C}_i$, is given by:

$$0 = D \triangle \hat{C}_i - \gamma \hat{C}_i + \eta(x_{i,S}) /V_c \tag{3.4}$$

which has the solution

$$\hat{C}_i(x_{j,S}) = \sum_{j=1}^{N} \frac{\eta(x_{j,S})}{4D(1+\frac{1}{\lambda})} \cdot \frac{1}{r_j} \cdot e^{\frac{-(r_j-R)}{R\lambda}} \tag{3.21}$$

taking the form of a sum of molecule m that was secreted by each cell j and has diffused to the location of cell i [198]. N is the number of cells within the population, r_j is the distance from the center of cell j to the surface of cell i, and $\lambda = \sqrt{D/\gamma} / R$ is the effective interaction distance. If we separate the contributions of cell i and all other cells j, $j \neq i$, we get:

$$\hat{C}_i(x_{i,S}, x_{j,S}) = \hat{C}_i^S(x_{i,S}) + \hat{C}_i^N(x_{j,S}) \tag{3.22}$$

where the contribution of cell i to $\hat{C}_i$ is

$$\hat{C}_i^S(x_{i,S}) = \frac{\eta(x_{i,S})}{4D(1+\frac{1}{\lambda})} \tag{3.23}$$

and the contribution of all neighbor cells j to $\hat{C}_i$ is

$$\hat{C}_i^N(x_{j,S}) = \sum_{j\neq i} \frac{\eta(x_{j,S})}{4D(1+\frac{1}{\lambda})} \cdot \frac{1}{r_j} \cdot e^{\frac{-(r_j-R)}{R\lambda}} . \quad (3.24)$$

To determine the boundaries of social behavior in our model, we consider a tissue in which R, D, γ, and r_j are constant and all cells j, except cell i, have the same value for $x_{j,S}$. That is, $x_{j,S} = 1$ or $x_{j,S} = 0 \ \ \forall\, j \ \neq i$. Thus, we can further simplify $\hat{C}_i^S(x_{i,S})$ and $\hat{C}_i^N(x_{j,S})$ to:

$$\hat{C}_i^S(x_{i,S}) = \beta \cdot \eta(x_{i,S}) \ , \quad (3.25)$$

and

$$\hat{C}_i^N(x_{j,S}) = \beta \cdot \eta(x_{j,S})\delta \quad (3.26)$$

where

$$\beta = (4D(1+\tfrac{1}{\lambda}))^{-1}, \quad (3.27)$$

and

$$\delta = \sum_{j\neq i} \beta \cdot \frac{1}{r_j} \cdot e^{\frac{-(r_j-R)}{R\lambda}}. \quad (3.28)$$

Note that both β and δ are constant due to the static position of all cells. Finally, the boundaries describing regions A1, A2, A3, S1, S2, and S3 are given by the receptor activation threshold value at which it equals the different combinations of self- and neighbor-generated signal, $\hat{C}_i^S(x_{i,S}), x_{i,S} \in \{0,1\}$ and $\hat{C}_i^N(x_{j,S}), x_{j,S} \in \{0,1\}$. This gives four boundary equations:

$$\theta_1 = \hat{C}_i^S(x_{i,S}=0) + \hat{C}_i^N(x_{j,S}=0) = \beta + \beta\delta \quad (3.29)$$

$$\theta_2 = \hat{C}_i^S(x_{i,S}=0) + \hat{C}_i^N(x_{j,S}=1) = \beta + \beta\alpha\delta \quad (3.30)$$

$$\theta_3 = \hat{C}_i^S(x_{i,S}=1) + \hat{C}_i^N(x_{j,S}=0) = \beta\alpha + \beta\delta \quad (3.31)$$

$$\theta_4 = \hat{C}_i^S(x_{i,S} = 1) + \hat{C}_i^N(x_{j,S} = 1) = \beta\alpha + \beta\alpha\delta \tag{3.32}$$

From this set of equations we can see that $\theta_1 < \theta_2 < \theta_4$ and $\theta_1 < \theta_3 < \theta_4$ will hold for any values of β, δ , N, and $\alpha > 1$.

To determine the relative positions of θ_2 and θ_3, we first find their intersection, given by:

$$\theta_2 = \theta_3$$

$$\beta + \beta\alpha\delta = \beta\alpha + \beta\delta$$

$$1 + \alpha\delta = \alpha + \delta$$

$$1 - \alpha = (1 - \alpha)\delta$$

$$1 = \delta$$

$$1 = \sum_{j \neq i} \beta \cdot \frac{1}{r_j} \cdot e^{\frac{-(r_j - R)}{R\lambda_I}} \tag{3.33}$$

where λ_I is the value of λ at which θ_2 and θ_3 intersect. If $\lambda > \lambda_I$then $\theta_2 < \theta_3$ and, similarly, if $\lambda < \lambda_I$then $\theta_2 > \theta_3$.

3.6 SUPPLEMENTAL FIGURES

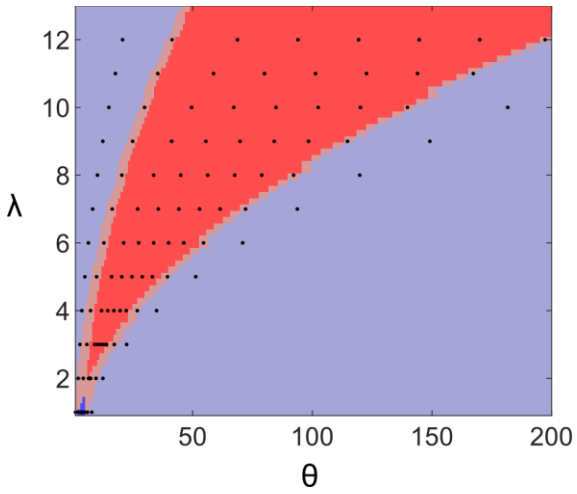


Figure S3.1. All sampled λ, θ values. For each of 12 λ values, 9 θ values were chosen - 1 in region A1; 1 in S1; 5 in S2 or A2; 1 in S3; and 1 in A3.

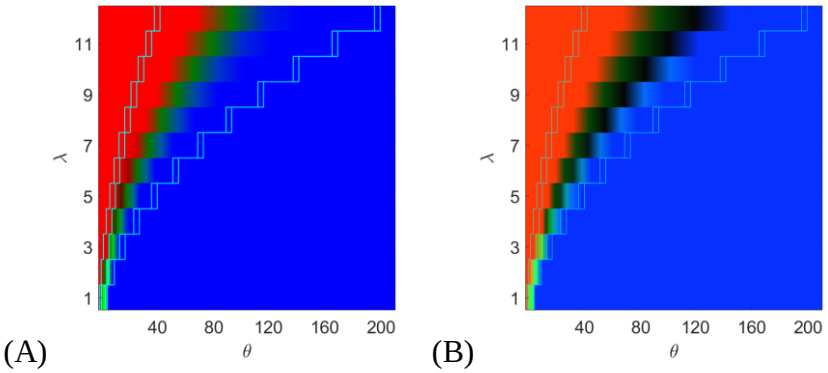


Figure S3.2. Example tissues in which asocial behavior is exhibited in social regions. **(A)** $D_H(A^{T_{\lambda,\theta}}, A^{T_{\lambda^{ON},\theta^{ON}}})$, $D_H(A^{T_{\lambda,\theta}}, A^{T_{\lambda^{OFF},\theta^{OFF}}})$, and $D_H(A^{T_{\lambda,\theta}}, A^{T_{\lambda^{SELF},\theta^{SELF}}})$ for a single tissue. The color is generated using (1-$D_H(A^{T_{\lambda,\theta}}, A^{T_{\lambda^{ON},\theta^{ON}}})$,1-$D_H(A^{T_{\lambda,\theta}}, A^{T_{\lambda^{SELF},\theta^{SELF}}})$, 1-$D_H(A^{T_{\lambda,\theta}}, A^{T_{\lambda^{OFF},\theta^{OFF}}})$) as (R,G,B) values. In bright regions, $b_{\lambda,\theta}$ is low and tissues behave similarly to in the asocial regions - color indicates which asocial region. In dark regions, $b_{\lambda,\theta}$ is high and tissues behave differently than in asocial regions. $b_{\lambda,\theta}$ = 0 in regions S1, S3 and on the left and right edges of S2. Cells on the left edge behave the same as in the ON region, while cells on the right edge behave the same as in the OFF region. **(B)** $D_{KL}(D_{\lambda,\theta,}D_{\lambda^{ON},\theta^{ON}})$, $D_{KL}(D_{\lambda,\theta,}D_{\lambda^{OFF},\theta^{OFF}})$, $D_{KL}(D_{\lambda,\theta,}D_{\lambda^{SELF},\theta^{SELF}})$ for a single tissue. The color is generated using (1-$D_{KL}(D_{\lambda,\theta,}D_{\lambda^{ON},\theta^{ON}})$, 1-$D_{KL}(D_{\lambda,\theta,}D_{\lambda^{OFF},\theta^{OFF}})$, 1-$D_{KL}(D_{\lambda,\theta,}D_{\lambda^{SELF},\theta^{SELF}})$) as (R,G,B) values. In bright regions $\widehat{D}_{KL}(\lambda,\theta)$is low and tissues behave similarly to in the asocial regions - color indicates which asocial region. In dark regions, $\widehat{D}_{KL}(\lambda,\theta)$ is high and tissues behave differently than in asocial regions. $\widehat{D}_{KL}(\lambda,\theta)$= 0 within S1, S2, and S3. Tissue samples in S1 and on the left edge of S2 have the same composition as in the ON region and tissue samples in S3 and right edge of S2 have the same composition as in the OFF region.

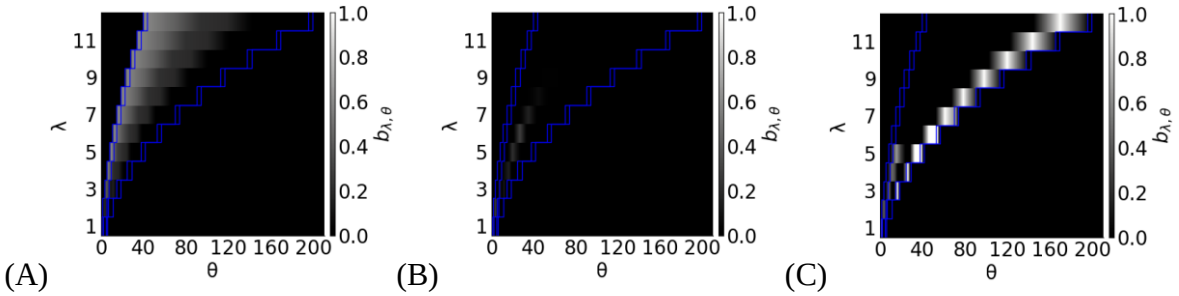


Figure S3.3. Three examples of variable behavior of $b_{\lambda,\theta}$ in different tissues.

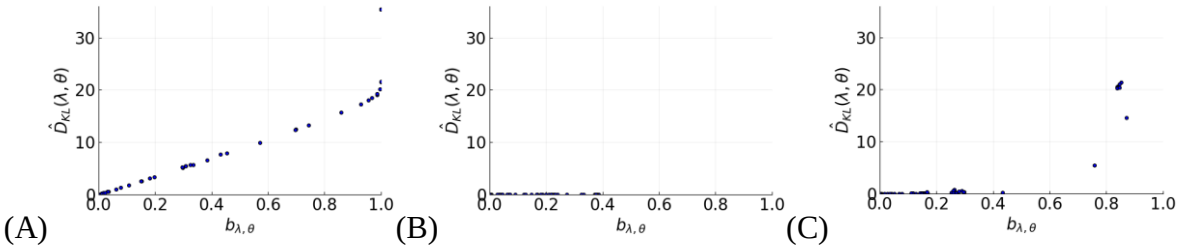


Figure S3.4. Three tissue examples of how changes in cellular behavior, $b_{\lambda,\theta}$ can relate to changes in tissue composition, $\widehat{D}_{KL}(\lambda,\theta)$. **(A)** A tissue with a strong positive relationship between $b_{\lambda,\theta}$ and $\widehat{D}_{KL}(\lambda,\theta)$. **(B)** A tissue with a very weak positive relationship between $b_{\lambda,\theta}$ and $\widehat{D}_{KL}(\lambda,\theta)$. **(C)** A tissue with a switch-like relationship between $b_{\lambda,\theta}$ and $\widehat{D}_{KL}(\lambda,\theta)$.

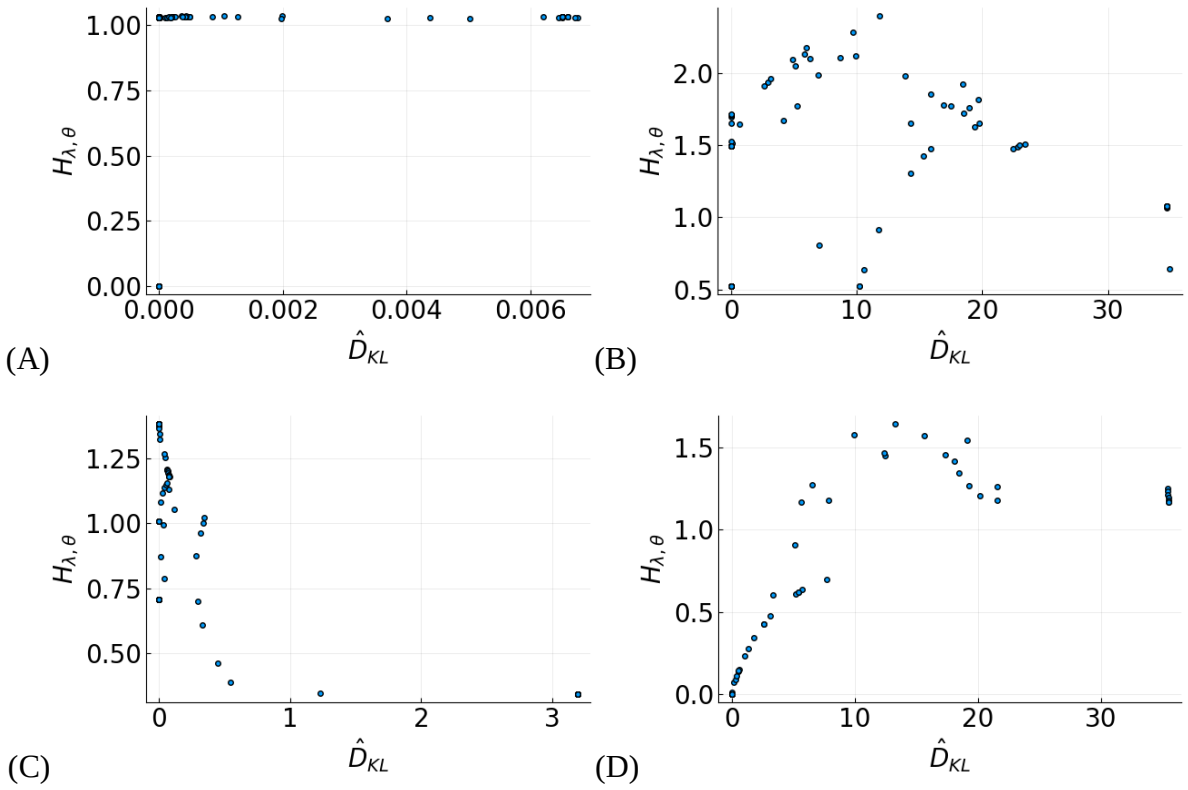


Figure S3.5. Relationships between $\widehat{D}_{KL}$ and $H_{\lambda,\theta}$. Each subplot shows the $\widehat{D}_{KL}(\lambda,\theta)$ and $H_{\lambda,\theta}$ for one tissue. **(A)** $H_{\lambda,\theta}$ does not change with $\widehat{D}_{KL}(\lambda,\theta)$. **(B)** $\widehat{D}_{KL}(\lambda,\theta)$ and $H_{\lambda,\theta}$ are mostly uncorrelated. **(C)** $\widehat{D}_{KL}(\lambda,\theta)$ and $H_{\lambda,\theta}$ are negatively correlated. **(D)** $\widehat{D}_{KL}(\lambda,\theta)$ and $H_{\lambda,\theta}$ are positively correlated.

Chapter 4. DISCUSSION

4.1 OVERVIEW

Many natural systems can be categorized as complex systems, with relatively simple components interacting to generate collective behaviors not easily predictable from the components themselves, like the flocking of birds or schooling of fish [3]. As such, a complex systems view is necessary in order to fully understand how behaviors of the population as a whole arise from interactions between individuals. In particular, our understanding of living systems, which are essentially nested complex systems, stands to benefit from the application of complex systems theory [204]. As the fundamental unit of life, cells provide a good starting point for developing an understanding of biological systems as complex systems.

In this thesis, I have explored cellular behavior using a complex systems approach, considering cells as both a system of intracellular components as well as a smaller unit within a larger cellular population. Specifically, I have focused on the cellular use of external signals, which can serve as a source of necessary information for cells as a system or as the interaction that binds cells together as a population. In order to study cells in these contexts, I took an abstract mechanistic modeling approach in which cellular dynamics are modeled using a Boolean network (BN) mathematical formalization. This approach allows for the relationship between agent behavior and population behavior to be explored without confounding factors of system-specific details. As a network of binary-valued nodes, Boolean networks are computationally and analytically tractable and have been shown to be capable of rich emergent behavior linked to cellular behavior [37,42,44,46].

4.2 Cells as Systems of Intracellular Components

In this thesis, I have focused on two system-level properties of cells as systems of intracellular components: (1) the capacity for information processing and (2) the system dynamics under externally driven input by time-varying signals.

To explore the capacity for information processing in cells, I use the reservoir computing framework. Reservoir computing has emerged in the last two decades [89,90] as a variant of neural networks that classifies inputs in real-time and does not involve training the internal topology of the system. The finding that fixed reservoirs are capable of signal processing has led researchers to show that many natural, fixed dynamical systems are capable of reservoir computing [136,205]. Thus, I chose to investigate cells, which are dynamical systems with fixed rules, as reservoirs in order to explore their information processing capacity. Though less common, there is some previous research showing the connection between cells and cellular populations to reservoir computers [128,130]. Importantly, these studies investigated specific systems as reservoir computers. I took an abstract approach in order to identify general trends in network properties and signal processing.

We found that large network size and input connectivity, both of which we associate with computational power, guarantee the flexibility to generate many different responses to the same signal with minimal retraining of the readout layer. However, there is significant heterogeneity in flexibility even with fewer nodes or less input connectivity, with a small subset of BNs being maximally flexible. This raises the question of whether there are structural properties that confer flexibility in these networks. Indeed we found that on average, networks with in-degree associated with critical behavior have higher average success across all tested tasks. However, we did not identify whether this feature is also associated with higher flexibility at low computational power.

Other structural differences, which are known to influence information processing capacity [61,86,206], could also be at play. Interestingly, regardless of flexibility, tasks that require a higher fidelity of memory of events that happened further in the past are the most difficult.

It would be valuable to gather data from real cellular systems and compare their networks as well as their signal processing requirements to our findings. If cells maintain high computational power then they retain flexibility in processing signals. Lower computational power would save resources, but would be less robust to changes in network structure since fewer such networks are highly flexible. However, if cells only require the ability to generate a small subset of responses, flexibility may be less important. Moreover, there may be a relationship between a cell's response network and the difficulty of its required signal responses. Do cells dedicate more resources when they are required to perform more difficult signal processing tasks or simply sacrifice accuracy?

Overall, the fact that our cellular models are capable of acting as reservoir computers has interesting implications for the adaptability of intracellular networks. Our work shows that a single network can be adapted to many different responses simply with a different readout. This suggests that the bulk of information processing can be performed by the same core network and then the output can be further processed by different sets of components in order to generate particular responses. Jones et al. [130] propose that a gene regulatory network acts as the reservoir and the readout for different responses is performed by specific proteins or DNA segments.

To explore the effects of time-varying signals on the dynamics of cells, we primarily used a network-of-networks modeling approach in which a population of cells described by networks interact through extracellular signals. Understanding the effects of external signals on the internal dynamics of BNs is still a relatively unexplored area, with no comprehensive study considering the dynamics of open BNs. We found that the behavior of a single BN can change drastically when

the external signal has direct or indirect feedback from the BN. The network can be forced into stable behavior it otherwise would not exhibit. This includes behavior that is not stable in closed BNs. Though we suspect that the dynamical regime of the BN likely changed as an open versus closed system, we did not perform the numerical experiments necessary to test this hypothesis. An analysis of how the system responds to perturbations could reveal any dynamical changes, though. We did observe a potential shift in dynamics in the BN reservoirs. As the reservoirs' connectivity to the input signal increased, the performance of networks with critical connectivity was reduced and those with sub-critical connectivity increased. Given that criticality is strongly associated with information processing capacity and that chaotic networks perform worse in our study, this suggests that the time-varying input is increasing the sensitivity of the network to perturbations. Since this topic is related to perturbation spreading, different structural and logical properties of Boolean networks likely have a role in how much a time-varying input influences the dynamics of the network [207,208].

4.3 Complex Cells as Interacting Units

In this thesis, I also investigated cells as the interacting agents within a cellular population. Rather than consider simple cellular automata cells, I modeled cells as complex networks that interact to form a larger complex system. Living systems are hierarchies of organization levels, each level considered as a complex system in its own right. For simplicity, most complex systems research only considers one organizational level at a time. In chapter 3, I added this complexity to examine the effects of cell-cell communication on cellular and population behavior. We found that communication between cellular networks leads to the emergence of novel cellular behavior and self-organization in tissue diversity. By tuning the parameters of communication, the number and

frequency of cellular behaviors as well as the degree of self-organization can be modified. Moreover, there may be a critical interaction distance that maximizes the effects of interaction. Interestingly, only a subset of cellular BNs exhibit unique behavior when communication occurs between cells. Those cellular BNs that do exhibit shifts in behavior due to cell-cell communication appear to separate into qualitatively different categories of response. Though we did not investigate it here, these findings could provide a predictable relationship between the lowest level of the system, the intracellular network, with the highest level of the system, the population behavior. Overall, our findings agree with the results from similar models by Damiani et al. [105] and Flann et al. [104].

Given that a significant portion of complex systems research considers one level of organization at a time, it is pertinent to ask whether more complex agents generate qualitatively different population behavior. Certainly, by accounting for two levels of organization, we were able to observe relationships that were non-obvious otherwise. Namely, we were able to identify the emergence of novel cellular behavior when signaling occurs between cells. It is possible that this could be observed in simple cellular automata models of cells by looking at the temporal dynamics of a binary-valued cell; however, the framework of cellular BNs makes the observation much more apparent and tangible. Furthermore, it is unclear how adding layers of complexity will actually change the system's qualitative behavior. In the work covered here, I only considered population level behaviors that would be observable in simple cellular automata, such as heterogeneity. Undoubtedly, more complex agents will generate more diverse behaviors, which can increase the relative complexity of the observed population-level properties. Ultimately, the question is whether we can learn more about a system by accounting for the multi-scale complexity. More studies with a network-of-networks approach would help to explore these ideas.

For example, the framework used here could be extended to examine different collective behaviors, such as synchronization or signal propagation (percolation). Computing in cellular populations of complex cells is another potential area of research. It has been established that populations of cells can act as reservoir computers. Our model could be extended to include cell-cell communication in order to provide insights into population level computing. Whatever the behavior being investigated, further work in multi-scale complex systems is warranted.

4.4 CONCLUSION

Using a complex systems framework, I was able to probe how cells can use external signals for their own benefit as well as for the benefit of the population by analyzing the behavior of cell-like agents. I identified relationships between the properties of the intracellular and cellular networks to collective properties of the simulated cells and cellular populations. To study these systems, I used two relatively novel modeling frameworks: Boolean network reservoir computers and 3D populations of interconnected Boolean networks. Using these methods, I was able to show that cells can be very flexible in generating diverse responses to external signals, but that introducing those signals can dramatically alter cellular behavior. These changes include both the class of dynamical behavior that the cell exhibits as well as the range of specific stable behaviors that a cell expresses. I also showed that cells can collectively utilize external signaling in the form of cell-cell communication to self-organize population diversity and cell type composition. Importantly, these findings are in systems that were not evolved or otherwise specifically constructed to exhibit these behaviors, indicating that they are simply a result of cell-environment and cell-cell interactions. Further work may reveal the role of cellular network structure on the effects of external signals on cellular behavior as well as population self-organization. Additional

analysis may also show other instances of self-organization via cell-cell communication, such as in generating spatial patterning.

BIBLIOGRAPHY

1. Lotka, A. J. Analytical Note on Certain Rhythmic Relations in Organic Systems. Proc. Natl. Acad. Sci. U. S. A. 6, 410–415 (1920).

2. Volterra, V. Fluctuations in the Abundance of a Species considered Mathematically. Nature 119, 12–13 (1927).

3. Bar-yam, Y. Dynamics of Complex Systems. (Westview Press, 2003).

4. Selinger, J. V. Ising Model for Ferromagnetism. In: Introduction to the Theory of Soft Matter: Soft and Biological Matter. 7–24 (Springer, 2016).

5. Steels, L. Language as a Complex Adaptive System. Parallel Problem Solving from Nature VI, 17–26 (2000).

6. Marcon, L. & Sharpe, J. Turing patterns in development: what about the horse part? Curr. Opin. Genet. Dev. 22, 578–584 (2012).

7. Turing, A. M. The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 237, 37–72 (1952).

8. Mitchell, M. Complex systems: Network thinking. Artificial Intelligence 170, 1194–1212 (2006).

9. Mitchell, M. Biological Computation. The Computer Journal 55, 852–855 (2012).

10. Lizier, J. T. The Local Information Dynamics of Distributed Computation in Complex Systems. Springer Theses (2013).

11. Newman, M. E. J. Resource Letter CS–1: Complex Systems. Am. J. Phys. 79, 800–810 (2011).

12. Wolf-Branigin, M. Introduction: The History and Theory of Complexity. In: Using Complexity Theory for Research and Program Evaluation. 1–31 (Oxford University Press, 2013).

13. Dougherty, E. R. & Bittner, M. L. Epistemology of the Cell: A Systems Perspective on Biological Knowledge. (John Wiley & Sons, 2011).

14. Van Regenmortel, M. H. V. Reductionism and complexity in molecular biology. EMBO reports 5, 1016–1020 (2004).

15. Mazzocchi, F. Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory. EMBO reports 9, 10–14 (2008).

16. Coveney, P. V., Dougherty, E. R. & Highfield, R. R. Big data need big theory too. Philos. Trans. A Math. Phys. Eng. Sci. 374, 20160153 (2016).

17. Phair, R. D. Mechanistic modeling confronts the complexity of molecular cell biology. Mol. Biol. Cell 25, 3494–3496 (2014).

18. Stelling, J. Mathematical models in microbial systems biology. Current Opinion in Microbiology 7, 513–518 (2004).

19. Palsson, B. The challenges of in silico biology. Nature Biotechnology 18, 1147–1150 (2000).

20. Weiss, J. N., Qu, Z. & Garfinkel, A. Understanding biological complexity: lessons from the past. FASEB J. 17, 1–6 (2003).

21. Chen, K. C. et al. Kinetic Analysis of a Molecular Model of the Budding Yeast Cell Cycle. Molecular Biology of the Cell 11, 369–391 (2000).

22. Weis, M. C., Avva, J., Jacobberger, J. W. & Sreenath, S. N. A Data-Driven, Mathematical Model of Mammalian Cell Cycle Regulation. PLoS ONE 9, e97130 (2014).

23. Sible, J. C. & Tyson, J. J. Mathematical modeling as a tool for investigating cell cycle control networks. Methods 41, 238–247 (2007).

24. Eftimie, R., Gillard, J. J. & Cantrell, D. A. Mathematical Models for Immunology: Current State of the Art and Future Research Directions. Bulletin of Mathematical Biology 78, 2091–2134 (2016).

25. Girel, S., Arpin, C., Marvel, J., Gandrillon, O. & Crauste, F. Model-based assessment of the Role of Uneven Partitioning of Molecular Content on Heterogeneity and Regulation of Differentiation in CD8 T-cell Immune Responses. Frontiers in Immunology 10, 230 (2019).

26. Pérez-Velázquez, J., Gölgeli, M. & García-Contreras, R. Mathematical Modelling of Bacterial Quorum Sensing: A Review. Bull. Math. Biol. 78, 1585–1639 (2016).

27. Marenda, M., Zanardo, M., Trovato, A., Seno, F. & Squartini, A. Modeling quorum sensing trade-offs between bacterial cell density and system extension from open boundaries. Sci. Rep. 6, 39142 (2016).

28. Chmielecki, J. et al. Optimization of dosing for EGFR-mutant non-small cell lung cancer with evolutionary cancer modeling. Sci. Transl. Med. 3, 90ra59 (2011).

29. Howard, G. R., Johnson, K. E., Rodriguez Ayala, A., Yankeelov, T. E. & Brock, A. A multi-state model of chemoresistance to characterize phenotypic dynamics in breast cancer. Sci. Rep. 8, 12058 (2018).

30. Durrett, R., Foo, J., Leder, K., Mayberry, J. & Michor, F. Intratumor Heterogeneity in Evolutionary Models of Tumor Progression. Genetics 188, 461–477 (2011).

31. Olimpio, E. P., Dang, Y. & Youk, H. Statistical Dynamics of Spatial-Order Formation by Communicating Cells. iScience 2, 27–40 (2018).

32. Calabretta, R. & Neirotti, J. Adaptive Agents in Changing Environments, the Role of Modularity. Neural Processing Letters 42, 257–274 (2015).

33. Zaman, L. et al. Coevolution drives the emergence of complex traits and promotes evolvability. PLoS Biol. 12, e1002023 (2014).

34. Kashtan, N. & Alon, U. Spontaneous evolution of modularity and network motifs. Proc. Natl. Acad. Sci. U. S. A. 102, 13773–13778 (2005).

35. Crutchfield, J. P. & Mitchell, M. The evolution of emergent computation. Proc. Natl. Acad. Sci. U. S. A. 92, 10742–10746 (1995).

36. Hein, A. M. et al. The evolution of distributed sensing and collective computation in animal populations. Elife 4, e10955 (2015).

37. Kauffman, S. Homeostasis and differentiation in random genetic control networks. Nature 224, 177–178 (1969).

38. Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. J. Theor. Biol. 128, 11–45 (1987).

39. Kauffman, S. A. & Smith, R. G. Adaptive automata based on Darwinian selection. Physica D: Nonlinear Phenomena 22, 68–82 (1986).

40. Bornholdt, S. Boolean network models of cellular regulation: prospects and limitations. J. R. Soc. Interface 5 Suppl 1, S85–94 (2008).

41. Shmulevich, I. & Dougherty, E. R. Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks. (SIAM, 2010).

42. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. Proc. Natl. Acad. Sci. U. S. A. 101, 4781–4786 (2004).

43. Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. Random Boolean network models and the yeast transcriptional network. Proc. Natl. Acad. Sci. U. S. A. 100, 14796–14799 (2003).

44. Albert, R. & Othmer, H. G. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. J. Theor. Biol. 223, 1–18 (2003).

45. Joo, J. I., Zhou, J. X., Huang, S. & Cho, K.H. Determining Relative Dynamic Stability of Cell States Using Boolean Network Model. Sci. Rep. 8, 12077 (2018).

46. Yachie-Kinoshita, A. et al. Modeling signaling-dependent pluripotency with Boolean logic to predict cell fate transitions. Molecular Systems Biology 14, e7952 (2018).

47. Green, D. G., Leishman, T. G. & Sadedin, S. The Emergence of Social Consensus in Boolean Networks. 2007 IEEE Symposium on Artificial Life (2007).

48. Alexander, J. M. & McKenzie Alexander, J. Random Boolean Networks and Evolutionary Game Theory. Philosophy of Science 70, 1289–1304 (2003).

49. Galstyan, A. & Lerman, K. Adaptive Boolean networks and minority games with time-dependent capacities. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 66, 015103 (2002).

50. Caetano, M. A. L. & Yoneyama, T. Boolean network representation of contagion dynamics during a financial crisis. Physica A: Statistical Mechanics and its Applications 417, 1–6 (2015).

51. Jumadinova, J., Matache, M. T. & Dasgupta, P. A Multi-agent Prediction Market Based on Boolean Network Evolution. 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (2011).

52. Roli, A., Manfroni, M., Pinciroli, C. & Birattari, M. On the Design of Boolean Network Robots. In: Di Chio C. et al. (eds) Applications of Evolutionary Computation. EvoApplications 2011. Lecture Notes in Computer Science, 6624 (Springer, 2011).

53. Roli, A. et al. Dynamical Properties of Artificially Evolved Boolean Network Robots. In: Gavanelli M., Lamma E., Riguzzi F. (eds) AI*IA 2015 Advances in Artificial Intelligence. AI*IA 2015. Lecture Notes in Computer Science, 9336 (Springer, 2015)

54. Garattoni, L., Roli, A., Amaducci, M., Pinciroli, C. & Birattari, M. Boolean Network Robotics as an Intermediate Step in the Synthesis of Finite State Machines for Robot Control. In: Advances in Artificial Life, ECAL 2013 (MIT Press, 2013).

55. Fath, B. D. & Grant, W. E. Ecosystems as evolutionary complex systems: Network analysis of fitness models. Environmental Modelling & Software 22, 693–700 (2007).

56. Campbell, C., Yang, S., Albert, R. & Shea, K. A network model for plant-pollinator community assembly. Proc. Natl. Acad. Sci. U. S. A. 108, 197–202 (2011).

57. Kauffman, S. A. Emergent properties in random complex automata. Physica D: Nonlinear Phenomena 10, 145–156 (1984).

58. Drossel, B. Random Boolean Networks. In: Heinz Georg Schuster (eds) Reviews of Nonlinear Dynamics and Complexity, 69–110 (Wiley-VCH, 2009).

59. Aldana, M., Coppersmith, S. & Kadanoff, L. P. Boolean Dynamics with Random Couplings. In: Kaplan E., Marsden J.E., Sreenivasan K.R. (eds) Perspectives and Problems in Nonlinear Science, 23–89 (Springer, 2003).

60. Mäki-Marttunen, T., Kesseli, J. & Nykter, M. Balance between Noise and Information Flow Maximizes Set Complexity of Network Dynamics. PLoS ONE 8, e56523 (2013).

61. Krawitz, P. & Shmulevich, I. Basin Entropy in Boolean Network Ensembles. Physical Review Letters 98, 158701 (2007).

62. Reichl, M. D. & Bassler, K. E. Canalization in the critical states of highly connected

networks of competing Boolean nodes. Physical Review E 84, 056103 (2011).

63. Correale, L., Leone, M., Pagnani, A., Weigt, M. & Zecchina, R. Core percolation and onset of complexity in boolean networks. Phys. Rev. Lett. 96, 018101 (2006).

64. Ribeiro, A. S., Kauffman, S. A., Lloyd-Price, J., Samuelsson, B. & Socolar, J. E. S. Mutual information in random Boolean models of regulatory networks. Physical Review E 77, 011901 (2008).

65. Darabos, C. et al. Additive functions in boolean models of gene regulatory network modules. PLoS One 6, e25110 (2011).

66. Shmulevich, I., Lähdesmäki, H., Dougherty, E. R., Astola, J. & Zhang, W. The role of certain Post classes in Boolean network models of genetic networks. Proc. Natl. Acad. Sci. U. S. A. 100, 10734–10739 (2003).

67. Szejka, A. & Drossel, B. Evolution of canalizing Boolean networks. The European Physical Journal B 56, 373–380 (2007).

68. Mihaljev, T. & Drossel, B. Evolution of a population of random Boolean networks. The European Physical Journal B 67, 259–267 (2009).

69. Oikonomou, P. & Cluzel, P. Effects of topology on network evolution. Nature Physics 2, 532–536 (2006).

70. Aldana, M., Balleza, E., Kauffman, S. & Resendiz, O. Robustness and evolvability in genetic regulatory networks. J. Theor. Biol. 245, 433–448 (2007).

71. Samuelsson, B. & Troein, C. Random maps and attractors in random Boolean networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 72, 046112 (2005).

72. Drossel, B. Number of attractors in random Boolean networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 72, 016110 (2005).

73. Kauffman, S. A. Origins of Order in Evolution: Self-Organization and Selection. In: Varela F.J., Dupuy JP. (eds) Understanding Origins. Boston Studies in the Philosophy and History of Science, 130. 153–181 (Springer1992).

74. Klemm, K. & Bornholdt, S. Stable and unstable attractors in Boolean networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 72, 055101 (2005).

75. Rämö, P., Kesseli, J. & Yli-Harja, O. Perturbation avalanches and criticality in gene regulatory networks. J. Theor. Biol. 242, 164–170 (2006).

76. Daniels, B. C. et al. Criticality Distinguishes the Ensemble of Biological Regulatory Networks. Phys. Rev. Lett. 121, 138102 (2018).

77. Mora, T. & Bialek, W. Are Biological Systems Poised at Criticality? Journal of Statistical Physics 144, 268–302 (2011).

78. Binney, J. The Theory of Critical Phenomena: An Introduction to the Renormalization Group. (Oxford University Press, 1992).

79. Solé, R. V., Manrubia, S. C., Luque, B., Delgado, J. & Bascompte, J. Phase transitions and complex systems: Simple, nonlinear models capture complex systems at the edge of chaos. Complexity 1, 13–26 (1996).

80. Muñoz, M. A. Colloquium: Criticality and dynamical scaling in living systems. Rev. Mod. Phys. 90, 031001 (2018).

81. Roli, A., Villani, M., Filisetti, A. & Serra, R. Dynamical Criticality: Overview and Open Questions. Journal of Systems Science and Complexity 31, 647–663 (2018).

82. Luque, B. & Solé, R. V. Lyapunov exponents in random Boolean networks. Physica A: Statistical Mechanics and its Applications 284, 33–45 (2000).

83. Shmulevich, I. & Kauffman, S. A. Activities and sensitivities in boolean network models. Phys. Rev. Lett. 93, 048701 (2004).

84. Nykter, M. et al. Critical networks exhibit maximal information diversity in structure-dynamics relationships. Phys. Rev. Lett. 100, 058702 (2008).

85. Goudarzi, A., Teuscher, C., Gulbahce, N. & Rohlf, T. Emergent Criticality through Adaptive Information Processing in Boolean Networks. Physical Review Letters 108, 128702 (2012).

86. Rämö, P., Kauffman, S., Kesseli, J. & Yli-Harja, O. Measures for information propagation in Boolean networks. Physica D: Nonlinear Phenomena 227, 100–104 (2007).

87. Niiranen, S., Ribeiro, A. (eds). Information Processing in Biological Systems. (Springter, 1985).

88. Tkačik, G. & Bialek, W. Information Processing in Living Systems. Annual Review of Condensed Matter Physics 7, 89–117 (2016).

89. Jaeger, H. Adaptive Nonlinear System Identification with Echo State Networks. In: Becker, S., Thrun S., Obermayer K. (eds). Advances in Neural Information Processing Systems, 15, 609–616 (MIT Press, 2003).

90. Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: a new framework for neural computation based on perturbations. Neural Comput. 14, 2531–2560 (2002).

91. Büsing, L., Schrauwen, B. & Legenstein, R. Connectivity, dynamics, and memory in reservoir computing with binary and analog neurons. Neural Comput. 22, 1272–1311 (2010).

92. Echlin, M., Aguilar, B., Notarangelo, M., Gibbs, D. L. & Shmulevich, I. Flexibility of Boolean Network Reservoir Computers in Approximating Arbitrary Recursive and Non-

recursive Binary Filters. Entropy 20, 954 (2018).

93. Villani, M., Barbieri, A. & Serra, R. A Dynamical Model of Genetic Networks for Cell Differentiation. PLoS ONE 6, e17703 (2011).

94. Wagner, A. Circuit topology and the evolution of robustness in two-gene circadian oscillators. Proc. Natl. Acad. Sci. U. S. A. 102, 11775–11780 (2005).

95. Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. Evolution of genetic redundancy. Nature 388, 167–171 (1997).

96. Friedman, N., Cai, L. & Xie, X. S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Phys. Rev. Lett. 97, 168302 (2006).

97. Nochomovitz, Y. D. & Li, H. Highly designable phenotypes and mutational buffers emerge from a systematic mapping between network topology and dynamic output. Proc. Natl. Acad. Sci. U. S. A. 103, 4180–4185 (2006).

98. Kaneko, K. & Yomo, T. Isologous diversification: a theory of cell differentiation. Bull. Math. Biol. 59, 139–196 (1997).

99. Solé, R. V. & Valverde, S. Spontaneous emergence of modularity in cellular networks. J. R. Soc. Interface 5, 129–133 (2008).

100. Gollo, L. L., Mirasso, C. & Eguíluz, V. M. Signal integration enhances the dynamic range in neuronal systems. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 85, 040902 (2012).

101. Larkin, J. W. et al. Signal Percolation within a Bacterial Community. Cell Syst 7, 137–145.e3 (2018).

102. Thurley, K., Wu, L. F. & Altschuler, S. J. Modeling Cell-to-Cell Communication Networks Using Response-Time Distributions. Cell Syst 6, 355–367.e5 (2018).

103. Youk, H. & Lim, W. A. Secreting and sensing the same molecule allows cells to achieve versatile social behaviors. Science 343, 1242782 (2014).

104. Flann, N. S., Mohamadlou, H. & Podgorski, G. J. Kolmogorov complexity of epithelial pattern formation: the role of regulatory network configuration. Biosystems. 112, 131–138 (2013).

105. Damiani, C., Kauffman, S. A., Villani, M., Colacci, A. & Serra, R. Cell–cell interaction and diversity of emergent behaviours. IET Systems Biology 5, 137–144 (2011).

106. Kitano, H. Systems biology: a brief overview. Science 295, 1662–1664 (2002).

107. Shivdasani, R. A. Limited gut cell repertoire for multiple hormones. Nat. Cell Biol. 20, 865–867 (2018).

108. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity.

Bull. Math. Biophys. 5, 115–133 (1943).

109. Lu, Z. et al. Reservoir observers: Model-free inference of unmeasured variables in chaotic systems. Chaos 27, 041102 (2017).

110. Pathak, J., Lu, Z., Hunt, R. B., Girvan, M. & Ott, E. Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data. Chaos 27, 121102 (2017).

111. Fonollosa, J., Sheik, S., Huerta, R. & Marco, S. Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring. Sens. Actuators B Chem. 215, 618-629 (2015).

112. Caluwaerts, K., D'Haene, M., Verstraeten, D. & Schrauwen, B. Locomotion without a brain: physical reservoir computing in tensegrity structures. Artif. Life 19, 35–66 (2013).

113. Aaser, P. et al. Towards making a cyborg: A closed-loop reservoir-neuro system. In: Advances in Artificial Life, ECAL 2017 (MIT Press, 2017).

114. Antonelo, A. E., Schrauwen, B. & Van Campenhout, J. Generative Modeling of Autonomous Robots and their Environments using Reservoir Computing. Neural Process. Letters 26, 233-249 (2007).

115. Bianchi, F. M., Santis, E. D., Rizzi, A. & Sadeghian, A. Short-Term Electric Load Forecasting Using Echo State Networks and PCA Decomposition. IEEE Access 3, 1931–1943 (2015).

116. Gallicchio, C., Micheli, A. A preliminary application of echo state networks to emotion recognition. In: Fourth International Workshop EVALITA (Pisa University Press, 2014)

117. Gallicchio, C. A Reservoir Computing Approach for Human Gesture Recognition from Kinect Data. In: Proceedings of the Second Italian Workshop on Artificial Intelligence for Ambient Assisted Living (AI*IA, 2016).

118. Waibel, A. Modular Construction of Time-Delay Neural Networks for Speech Recognition. Neural Comput. 1, 39–46 (1989).

119. Triefenbach, F., Jalalvand, A., Schrauwen, B. & Martens, J.P. Phoneme Recognition with Large Hierarchical Reservoirs. In: Lafferty, J.D., Williams, C.K.I., Shawe-Taylor, J., Zemel, R.S., Culotta, A. (eds) Advances in Neural Information Processing Systems 23, 2307–2315 (Curran Associates, Inc., 2010).

120. Palumbo, F., Gallicchio, C., Pucci, R. & Micheli, A. Human activity recognition using multisensor data fusion based on Reservoir Computing. J. Ambient Intell. Smart Environ. 8, 87-107 (2016)

121. Luz, E. J. da S., Schwartz, W. R., Cámara-Chávez, G. & Menotti, D. ECG-based heartbeat classification for arrhythmia detection: A survey. Comput. Methods Programs Biomed. 127, 144–164 (2016).

122. Merkel, C., Saleh, Q., Donahue, C. & Kudithipudi, D. Memristive Reservoir Computing Architecture for Epileptic Seizure Detection. Procedia Comput. Sci. 41, 249–254 (2014).

123. Buteneers, P. et al. Automatic detection of epileptic seizures on the intra-cranial electroencephalogram of rats using reservoir computing. Artif. Intell. Med. 53, 215–223 (2011).

124. Ayyagari, S. Reservoir computing approaches to EEG-based detection of microsleeps. Ph.D. Thesis (University of Canterbury, 2017).

125. Kainz, P., Burgsteiner, H., Asslaber, M. & Ahammer, H. Robust Bone Marrow Cell Discrimination by Rotation-Invariant Training of Multi-class Echo State Networks. In: Iliadis L., Jayne C. (eds) Engineering Applications of Neural Networks. Communications in Computer and Information Science. 517, 390–400 (Springer, 2015).

126. Reid, D. & Barrett-Baxendale, M. Glial Reservoir Computing. In: Second UKSIM European Symposium on Computer Modeling and Simulation. 81–86 (IEEE Computer Society, 2008).

127. Enel, P., Procyk, E., Quilodran, R. & Dominey, P. F. Reservoir Computing Properties of Neural Dynamics in Prefrontal Cortex. PLoS Comput. Biol. 12, e1004967 (2016).

128. Yamazaki, T. & Tanaka, S. The cerebellum as a liquid state machine. Neural Netw. 20, 290–297 (2007).

129. Dai, X. Genetic Regulatory Systems Modeled by Recurrent Neural Network. In: Advances in Neural Networks, Proceedings of the International Symposium on Neural Networks. Lecture Notes in Computer Science. 3174, 519–524 (Springer, 2004).

130. Jones, B., Stekel, D., Rowe, J. & Fernando, C. Is there a Liquid State Machine in the Bacterium Escherichia Coli? 2007 IEEE Symposium on Artificial Life (IEEE, 2007)

131. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. J. R. Stat. Soc. Series B Stat. Methodol. 58, 267–288 (1996).

132. Lynn, P. A. Recursive digital filters for biological signals. Med. Biol. Eng. 9, 37–43 (1971).

133. Burian, A. & Kuosmanen, P. Tuning the smoothness of the recursive median filter. IEEE Trans. Signal Process. 50, 1631–1639 (2002).

134. Shmulevich, I., Yli-Harja, O., Egiazarian, K. & Astola, J. Output distributions of recursive stack filters. IEEE Signal Process. Lett. 6, 175–178 (1999).

135. Dambre, J., Verstraeten, D., Schrauwen, B. & Massar, S. Information processing capacity of dynamical systems. Sci. Rep. 2, 514 (2012).

136. Fernando, C. & Sojakka, S. Pattern Recognition in a Bucket. In: Banzhaf W., Ziegler J., Christaller T., Dittrich P., Kim J.T. (eds) Advances in Artificial Life. ECAL 2003. Lecture Notes in Computer Science. 2801, 588–597 (Springer, 2003).

137. Kulkarni, S. M. & Teuscher, C. Memristor-based reservoir computing. In: 2012 IEEE ACM International Symposium on Nanoscale Architectures. (IEEE, 2012).

138. Dale, M., Miller, J. F., Stepney, S. & Trefzer, M. A. Evolving Carbon Nanotube Reservoir Computers. In: Proceedings of the UCNC 2016: Unconventional and Natural Computation. (Springer International Publishing, 2016).

139. Kauffman, S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. 22, 437–467 (1969).

140. Snyder, D., Goudarzi, A., Teuscher, C. Finding optimal random boolean networks for reservoir computing. In: Proceedings of the Thirteenth International Conference on the Simulation and Synthesis of Living Systems. (MIT Press, 2009).

141. Derrida, B. & Pomeau, Y. Random networks of automata: a simple annealed approximation. EPL 1, 45 (1986).

142. Van der Sande, G., Brunner, D. & Soriano, M. C. Advances in photonic reservoir computing. Nanophotonics 6, 8672 (2017).

143. Shmulevich, I., Dougherty, E. R. & Zhang, W. From Boolean to probabilistic Boolean networks as models of genetic regulatory networks. Proc. IEEE 90, 1778–1792 (2002).

144. de Jong, H. Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. J. Comput. Biol. 9, 67–103 (2002).

145. Davidich, M. I. & Bornholdt, S. Boolean network model predicts cell cycle sequence of fission yeast. PLoS One 3, e1672 (2008).

146. Fumiã, H. F. & Martins, M. L. Boolean network model for cancer pathways: predicting carcinogenesis and targeted therapy outcomes. PLoS One 8, e69008 (2013).

147. Serra, R., Villani, M., Barbieri, A., Kauffman, S. A. & Colacci, A. On the dynamics of random Boolean networks subject to noise: attractors, ergodic sets and cell types. J. Theor. Biol. 265, 185–193 (2010).

148. Helikar, T., Konvalina, J., Heidel, J. & Rogers, J. A. Emergent decision-making in biological signal transduction networks. Proc. Natl. Acad. Sci. U. S. A. 105, 1913–1918 (2008).

149. Thakar, J., Pilione, M., Kirimanjeswara, G., Harvill, E. T. & Albert, R. Modeling systems-level regulation of host immune responses. PLoS Comput. Biol. 3, e109 (2007).

150. Snyder, D., Goudarzi, A. & Teuscher, C. Computational capabilities of random automata networks for reservoir computing. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 87, 042808 (2013).

151. Bertschinger, N. & Natschläger, T. Real-time computation at the edge of chaos in recurrent neural networks. Neural Comput. 16, 1413–1436 (2004).

152. Balleza, E. et al. Critical dynamics in genetic regulatory networks: examples from four kingdoms. PLoS One 3, e2456 (2008).

153. Torres-Sosa, C., Huang, S. & Aldana, M. Criticality is an emergent property of genetic networks that exhibit evolvability. PLoS Comput. Biol. 8, e1002669 (2012).

154. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011).

155. Cook, S., Dwork, C. & Reischuk, R. Upper and Lower Time Bounds for Parallel Random Access Machines without Simultaneous Writes. SIAM J. Comput. 15, 87–97 (1986).

156. Kahn, J., Kalai, G. & Linial, N. The Influence of Variables on Boolean Functions. In: Proceedings of the 29th Annual Symposium on Foundations of Computer Science (IEEE Computer Society, 1988).

157. Prindle, A. et al. Ion channels enable electrical communication in bacterial communities. Nature 527, 59–63 (2015).

158. Rutherford, S. T. & Bassler, B. L. Bacterial quorum sensing: its role in virulence and possibilities for its control. Cold Spring Harb. Perspect. Med. 2, (2012).

159. Williams, P. et al. Quorum sensing and the population-dependent control of virulence. Philos. Trans. R. Soc. Lond. B Biol. Sci. 355, 667–680 (2000).

160. Long, Z., Quaife, B., Salman, H. & Oltvai, Z. N. Cell-cell communication enhances bacterial chemotaxis toward external attractants. Sci. Rep. 7, 12855 (2017).

161. Bischofs, I. B., Hug, J. A., Liu, A. W., Wolf, D. M. & Arkin, A. P. Complexity in bacterial cell-cell communication: quorum signal integration and subpopulation signaling in the Bacillus subtilis phosphorelay. Proc. Natl. Acad. Sci. U. S. A. 106, 6459–6464 (2009).

162. Lopez, D., Vlamakis, H., Losick, R. & Kolter, R. Paracrine signaling in a bacterium. Genes Dev. 23, 1631–1638 (2009).

163. van Gestel, J., Vlamakis, H. & Kolter, R. From cell differentiation to cell collectives: Bacillus subtilis uses division of labor to migrate. PLoS Biol. 13, e1002141 (2015).

164. Yoon, H. S. & Golden, J. W. Heterocyst pattern formation controlled by a diffusible peptide. Science 282, 935–938 (1998).

165. Callahan, S. M. & Buikema, W. J. The role of HetN in maintenance of the heterocyst pattern in Anabaena sp. PCC 7120. Mol. Microbiol. 40, 941–950 (2001).

166. Vlamakis, H., Aguilar, C., Losick, R. & Kolter, R. Control of cell fate by the formation of an architecturally complex bacterial community. Genes Dev. 22, 945–953 (2008).

167. Ben-Jacob, E. et al. Cooperative formation of chiral patterns during growth of bacterial colonies. Phys. Rev. Lett. 75, 2899–2902 (1995).

168. van Gestel, J., Vlamakis, H. & Kolter, R. Division of Labor in Biofilms: the Ecology of Cell Differentiation. Microbiol Spectr 3, MB–0002–2014 (2015).

169. Niklas, K. J. & Newman, S. A. The origins of multicellular organisms. Evolution & Development 15, 41–52 (2013).

170. Rokas, A. The origins of multicellularity and the early history of the genetic toolkit for animal development. Annu. Rev. Genet. 42, 235–251 (2008).

171. Tameshige, T., Hirakawa, Y., Torii, K. U. & Uchida, N. Cell walls as a stage for intercellular communication regulating shoot meristem development. Front. Plant Sci. 6, 324 (2015).

172. Schumacher, J. A. et al. Intercellular calcium signaling in a gap junction-coupled cell network establishes asymmetric neuronal fates in C. elegans. Development 139, 4191–4201 (2012).

173. Haas, P. & Gilmour, D. Chemokine signaling mediates self-organizing tissue migration in the zebrafish lateral line. Dev. Cell 10, 673–680 (2006).

174. Ellison, D. et al. Cell–cell communication enhances the capacity of cell ensembles to sense shallow gradients during morphogenesis. Proceedings of the National Academy of Sciences 113, E679–E688 (2016).

175. Montero, J. A. et al. Role of FGFs in the control of programmed cell death during limb development. Development 128, 2075–2084 (2001).

176. Lohmann, I., McGinnis, N., Bodmer, M. & McGinnis, W. The Drosophila Hox gene deformed sculpts head morphology via direct regulation of the apoptosis activator reaper. Cell 110, 457–466 (2002).

177. Hart, Y. et al. Paradoxical signaling by a secreted molecule leads to homeostasis of cell levels. Cell 158, 1022–1032 (2014).

178. Zhou, X. et al. Circuit Design Features of a Stable Two-Cell System. Cell 172, 744–757.e17 (2018).

179. Wilson, S. E., Netto, M. & Ambrósio, R. Corneal cells: chatty in development, homeostasis, wound healing, and disease. American Journal of Ophthalmology 136, 530–536 (2003).

180. Nakahama, K.-I. Cellular communications in bone homeostasis and repair. Cell. Mol. Life Sci. 67, 4001–4009 (2010).

181. Gurtner, G. C., Werner, S., Barrandon, Y. & Longaker, M. T. Wound repair and regeneration. Nature 453, 314–321 (2008).

182. Kobayashi, M. & Shigenaka, Y. The mode of action of acetylcholine and 5-hydroxytryptamine at the neuromuscular junctions in a molluscan muscle (radular protractor). Comparative Biochemistry and Physiology Part C: Comparative Pharmacology

60, 115–122 (1978).

183. Zimmerman, L. M., Vogel, L. A. & Bowden, R. M. Understanding the vertebrate immune system: insights from the reptilian perspective. J. Exp. Biol. 213, 661–671 (2010).

184. Shklarsh, A., Ariel, G., Schneidman, E. & Ben-Jacob, E. Smart swarms of bacteria-inspired agents with performance adaptable interactions. PLoS Comput. Biol. 7, e1002177 (2011).

185. Payne, S. et al. Temporal control of self-organized pattern formation without morphogen gradients in bacteria. Mol. Syst. Biol. 9, 697 (2013).

186. Toda, S., Blauch, L. R., Tang, S. K. Y., Morsut, L. & Lim, W. A. Programming self-organizing multicellular structures with synthetic cell-cell signaling. Science 361, 156–162 (2018).

187. Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. Nature 469, 212–215 (2011).

188. Zuroff, T. R. & Curtis, W. R. Developing symbiotic consortia for lignocellulosic biofuel production. Appl. Microbiol. Biotechnol. 93, 1423–1435 (2012).

189. Hennig, S., Rödel, G. & Ostermann, K. Artificial cell-cell communication as an emerging tool in synthetic biology applications. J. Biol. Eng. 9, 13 (2015).

190. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. Nature 467, 167–173 (2010).

191. Smet, I. D., De Smet, I. & Beeckman, T. Asymmetric cell division in land plants and algae: the driving force for differentiation. Nature Reviews Molecular Cell Biology 12, 177–188 (2011).

192. Naldi, A., Carneiro, J., Chaouiya, C. & Thieffry, D. Diversity and plasticity of Th cell types predicted from regulatory network modelling. PLoS Comput. Biol. 6, e1000912 (2010).

193. Singh, A. M. et al. Cell-Cycle Control of Developmentally Regulated Transcription Factors Accounts for Heterogeneity in Human Pluripotent Cells. Stem Cell Reports 2, 398 (2014).

194. Revell, C., Blumenfeld, R. & Chalut, K. Force-based three-dimensional model predicts mechanical drivers of cell sorting. Proc. Biol. Sci. 286, 20182495 (2019)

195. Rudge, T. J., Federici, F., Steiner, P. J., Kan, A. & Haseloff, J. Cell polarity-driven instability generates self-organized, fractal patterning of cell layers. ACS Synth. Biol. 2, 705–714 (2013).

196. Boyer, M. & Wisniewski-Dyé, F. Cell-cell signaling in bacteria: not simply a matter of quorum. FEMS Microbiology Ecology 70, 1–19 (2009).

197. Mitri, S., Clarke, E. & Foster, K. R. Resource limitation drives spatial organization in microbial groups. ISME J. 10, 1471–1482 (2016).

198. Maire, T. & Youk, H. Molecular-Level Tuning of Cellular Autonomy Controls the Collective Behaviors of Cell Populations. Cell Syst 1, 349–360 (2015).

199. Berg, H. C. Random Walks in Biology. (Princeton University Press, 2018).

200. Kang, S., Kahan, S., McDermott, J., Flann, N. & Shmulevich, I. Biocellion: accelerating computer simulation of multicellular biological system models. Bioinformatics 30, 3101–3108 (2014).

201. Jin, M. et al. Yeast dynamically modify their environment to achieve better mating efficiency. Sci. Signal. 4, ra54 (2011).

202. Wartlick, O. et al. Dynamics of Dpp signaling and proliferation control. Science 331, 1154–1159 (2011).

203. Joncker, N. T., Fernandez, N. C., Treiner, E., Vivier, E. & Raulet, D. H. NK cell responsiveness is tuned commensurate with the number of inhibitory receptors for self-MHC class I: the rheostat model. J. Immunol. 182, 4572–4580 (2009).

204. Mazzocchi, F. Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory. EMBO Rep. 9, 10–14 (2008).

205. Dale, M., Miller, J. F., Stepney, S. & Trefzer, M. A. Evolving Carbon Nanotube Reservoir Computers. Unconventional Computation and Natural Computation 49–61 (2016).

206. Lizier, J. T., Pritam, S. & Prokopenko, M. Information dynamics in small-world Boolean networks. Artif. Life 17, 293–314 (2011).

207. Lu, Q. & Teuscher, C. Damage spreading in spatial and small-world random Boolean networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 89, 022806 (2014).

208. Fretter, C., Szejka, A. & Drossel, B. Perturbation propagation in random and evolved Boolean networks. New Journal of Physics 11, 033005 (2009).

VITA

Moriah L. Echlin

Contact Information
Email: moriah.echlin@gmail.com

Education

	Year	*Degree/Institution*	*Discipline*
Undergraduate	2011	B.Sc., Colorado State University	Biology, Mathematics(minor) GPA: 3.96/4.0
Doctorate	2019	Ph.D., University of Washington	Molecular/Cellular Biology GPA: 3.9/4.0

Research Experience

Apr. 2014 – June 2019 *Doctoral Thesis*
Laboratory of Dr. Ilya Shmulevich, Institute for Systems Biology
Researching the effects of cell-cell communication on behavior at the cellular and population level, including the generation of novel cellular dynamics, maintenance of tissue diversity, and coordination of complex functions.

Mar. 2012 – Dec. 2012 *Lab Technician*
Laboratory of Dr. Ruth Hufbauer, Department of Bioagriculture, Colorado State University
Researching the importance of genetic diversity and demography in the success of founding populations using a Tribolium castaneum model.

Aug. 2012 – Dec 2012 *Lab Technician*
Material Safety Lab, Hewlett-Packard, Fort Collins, CO
Analyzing electronic components for Restriction of Hazardous Substances guideline compliance.

Aug. 2009 – Dec. 2011 *Undergraduate Thesis*
Laboratory of Dr. Don Mykles, Department of Biology, Colorado State University
Researching the molecular signaling pathways that regulate crustacean molting with a focus on the role of TGF-β and TGF-β inhibitors.

Jan. 2009 – Sept. 2011 *Research Intern*
FEScUE Program under Drs. Michael Antolin and Simon Tavener, Colorado State University
Developing and utilizing new mathematical models for investigating disease-host interactions and disease management strategies in wildlife populations.

Publications

Echlin, M., Aguilar, B., Notarangelo, M., Gibbs, D. L. & Shmulevich, I. *Flexibility of Boolean Network Reservoir Computers in Approximating Arbitrary Recursive and Non-recursive Binary Filters.* Entropy 20, 954 (2018).

Funding

Completed

DGE-1256082, *NSF June 2014 – June 2017*

Title: Identifying mechanisms of stability in polymicrobial communities

Role: PI

The main goal of this project was to identify stable network architectures in microbial communities of three to five species using in silico methodology followed by empirical testing of those networks with bioengineered communities.

Presentations

2019 **M. Echlin**. *A Complex Systems Approach to Understanding Cells as Systems and Agents*. Thesis Defense, Seattle, WA.

2015 **M. Echlin**, I. Shmulevich. *Dynamics and Properties of Biological Networks*. The Institute for Systems Biology Annual Retreat, Leavenworth, WA.

2012 **M. Echlin**, D. Mykles. *Cloning and Characterization of a Novel Transforming Growth Factor β in Crustaceans*. Society for Integrative and Comparative Biology Conference, Charleston, SC.

2011 L. Chase, E. Costain, J. Drendel, **M. Echlin**, D. Hopkins, L. Jeffers, M. Mikucki, E. Moses, D. Neavin, S. Parks, K. Strand, R. Thornton, L. Wooten, M. Antolin, S. Field, S. Tavener. *An Evolutionary Stage-structured Disease Model*. CURC Symposium, Fort Collins, CO.

2010 E. Costain, J. Drendel, **M. Echlin**, D. Hopkins, M. Mikucki, S. Parks, R. Thornton, M. Antolin, S. Field, S. Tavener. *Spatial Modeling of Whitebark Pine and the Effects of Blister Rust*. CURC Symposium, Fort Collins, CO.

2009 **M. Echlin**, D. Hopkins, M. Antolin, S. Tavener. *Modeling of white pine blister rust*. National Institute for Mathematical and Biological Synthesis Undergraduate Conference, Knoxville, TN.

Teaching Experience

June 2018 – Aug. 2018

Undergraduate Researcher Mentor, Institute for Systems Biology

Designing a small research project for an undergraduate student and providing guidance through the research process over a 10 week internship.

Sept. 2014 – Dec. 2014

Teaching assistant, Foundations in Molecular Cell Biology, University of Washington

Providing classroom support for active learning activities, leading separate supplemental learning periods and grading.

Nov. 2014

Volunteer, Life Science Research Weekend, Pacific Science Center

Explaining systems biology concepts through interactive displays and activities to event participants (all ages).

July 2014
Teaching assistant, BioQuest, Center for Infectious Disease Research
Introducing high school seniors to basic infectious disease biology and laboratory techniques.

Other Professional Experience

2019 *Organizer*. Biomedical Hooding Ceremony, University of Washington

2016 *Organizer*. Molecular and Cellular Biology Program Symposium, Evolution: Making Sense of Biology

Academic and Professional Honors

NSF Graduate Fellow	National Science Foundation, GRFP	2014
Graduated Magna Cum Laude	Colorado State University	2011