The Physics of Open Ended Evolution

by

Alyssa M Adams

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2017 by the
Graduate Supervisory Committee:

Sara I Walker, Chair
Paul CW Davies
Theodore P Pavlic
Ralph V Chamberlin

ARIZONA STATE UNIVERSITY

December 2017

ABSTRACT

What makes living systems different than non-living ones? Unfortunately this question is impossible to answer, at least currently. Instead, we must face computationally tangible questions based on our current understanding of physics, computation, information, and biology. Yet we have few insights into how living systems might *quantifiably* differ from their non-living counterparts, as in a mathematical foundation to explain away our observations of biological evolution, emergence, innovation, and organization. The development of a theory of living systems, if at all possible, demands a mathematical understanding of how data generated by complex biological systems changes over time. In addition, this theory ought to be broad enough as to not be constrained to an Earth-based biochemistry. In this dissertation, the philosophy of studying living systems from the perspective of traditional physics is first explored as a motivating discussion for subsequent research. Traditionally, we have often thought of the physical world from a bottom-up approach: things happening on a smaller scale aggregate into things happening on a larger scale. In addition, the laws of physics are generally considered static over time. Research suggests that biological evolution may follow dynamic laws that (at least in part) change as a function of the state of the system. Of the three featured research projects, cellular automata (CA) are used as a model to study certain aspects of living systems in two of them. These aspects include self-reference, open-ended evolution, local physical universality, subjectivity, and information processing. Open-ended evolution and local physical universality are attributed to the vast amount of innovation observed throughout biological evolution. Biological systems may distinguish themselves in terms of information processing and storage, not outside the theory of computation. The final research project concretely explores real-world phenomenon by means of mapping dominance hierarchies in the

evolution of video game strategies. Though the main question of how life differs from non-life remains unanswered, the mechanisms behind open-ended evolution and physical universality are revealed.

DEDICATION

Dedicated to people who hate science. And to whoever has thought the odds of life emerging from randomness are the same as a twister randomly throwing together an airbus by tearing through a junkyard.

A big part of my time here at Arizona State was spent with Sundial, a student organization that fosters a positive community within the sciences (11 credit hours, to be exact). I've decided to carry many of those community-building aspects with me through the way I present science to others. As a result, I want my science to be accessible to everyone, including this dissertation. Not only does this dissertation showcase my accomplishments in research, but it also showcases the skills I've developed in Sundial's leadership and education programs. My goal is to make this dissertation as accessible as possible without sacrificing scientific rigor.

ACKNOWLEDGMENTS

How could I possibly recognize all the people in my life who have contributed to my this dissertation? At the risk of forgetting countless interactions, special thanks to (in order of appearance):

My mom for showing me the love of learning, my dad for showing me how to be creative, my Grandma and Papa for showing me the love of biology, my brother for showing me how to be chill, Mrs. White for showing me the love of mathematics, Mr. Arner for saying "If you want to be a physicist, you'd be really good at it" (I wouldn't have gone into physics without those words), John for showing me how to be resilient, Mr. James K Grimalkin for showing me how to be distracted, Riot Games for Teemo, Sara for showing me the vast lands of research, all members and visitors in ELIFE for the stimulating discussions and existential crises, The Templeton World Charity Foundation for the fellowship that funded my Research Assistanceship, Angelica for being an incredible hard worker and enthusiastic cheerleader throughout this dissertation process, ASU's NASA Space Grant Program for funding undergraduate research on cellular automata research, and the people who first thought of boiling coffee beans to make a beverage (the true unsung heroes of academia).

TABLE OF CONTENTS

CHAPTER Page

APPENDIX Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Chapter 1

THE ROAD MAP

1.1 PhD: Emphasis on Philosophy

Although finding reasonable answers to terribly difficult questions is the core challenge of science, it is quietly accompanied by the super important task of forming terribly difficult questions to answer. This includes things like identifying the right from of reference to ask the question from, or identifying what knowledge is needed in order to answer the question. It's also important to understand what aspect the question leaves out or what biases the question invokes.

The question that motivates this dissertation is "What makes living systems different than non-living ones?" This question cannot be answered directly as-is because of its biases, required prerequisite knowledge, and possibly because of the unknown aspects of life that the question doesn't cover. In order to answer this question eventually, answerable questions must precede it (along with some answers of course). The whole purpose of this dissertation is to attempt to answer one of these preceding questions.

So how is an answerable question formed? Why not the original question of "What makes living systems different than non-living ones?" Well if that were our starting point, then how can you, as a living being, begin to think about a world where life has never existed at all? Maybe imagining a rocky grey planet isn't all that difficult, but is it possible to identify parts of Earth that haven't been mangled by life? There

is a vast number of planets and not all of them are grey and rocky so it would be important to know what a planet completely devoid of life looks (and acts?) like.

Perhaps it would be easier to understand what life *is* instead of what life *isn't*. Suppose that we were able to define every living and dead organism here on Earth, it is still unclear how it would help us identify life in other parts of the universe. After all, it might not even be helpful in detecting life on planets where water and oxgen aren't present. The best we could hope for is an Earth-clone that contains all the same chemistry as we know and love and look for aliens that use a similar biology. Given that we don't have any idea of what other forms life is capable of taking, the odds of this happening are between 0% and 100%. It's anyone's guess and it all depends on what one is willing to call "life" and in what environment[1]. Philosophically, this is not unlike Plato's Allegory of the Cave (Plato 1943) which explores the lack of education and our ability to perceive the world. In this allegory, prisoners are chained to a wall in a cave and have only ever seen shadows cast on the wall from figures on a wall (Figure 1). Knowledge frees us from the cave and transforms our understanding of reality. We recognize the shadows are not reality, but a manufactured version of it. In the context of this dissertation, our understanding of living systems is limited by our subjective understanding of what life is.

Would it be possible to generalize life beyond living systems here on Earth? In the end, we ought to come up with a theory for life that is so broad, we wouldn't have to worry about life that doesn't use our same chemistry. Imagine if, in 100 years or so, we send scientists to study one of the planets in Tabby's star (the star system with the mysteriously dimming star, which is popularly thought to be a Dyson sphere

[1]For example, viruses are usually considered alive when procreating with a cell and not alive when not (Villarreal 2008)

Figure 1: *Antrum Platonicum*, or The Cave of Plato by Cornelis Cornelisz. van Haarlem (printed by Jan Saenredam in 1604). In this engraving, two groups of philosophers are separated by a wall. On top of the wall is a row of shadow-puppet figures, including Cupid and Bacchus. Both groups of philosophers debate from a different perspective (Haarlem 1604).

("TRAPPIST-1" 2017)). The spaceship lands in a whirl of steam and dust. Wearily, humans step off the ship onto the alien planet. There is no oxygen or water, yet silicon is everywhere. Just below the surface, supercritical CO_2 rivers gurgle between super-pressurized layers on crust (see Magliocco, Glaser, and Kneafsey 2013 for more on super-critical CO_2). Is there life on this planet? We must know what life is like independent of its manifestation here on Earth if we wish to know about living systems outside of Earth. Could we describe all of life in a few simple mathematical equations that abstract away the messy chemistry and environmental factors? Whether or not,

the important thing is that each idea that is encountered must be carefully scrutinized before moving forward. Otherwise the question might not be answerable, as in the case of the main motivating question.

Therefore, the rest of this dissertation is dedicated to answering smaller, more bite-sized questions that are carefully engineered to prod the "Big Question": *What makes living systems different than non-living ones?* The rest of this chapter (the biggest chapter) explores some smaller questions to ask and which ones to avoid asking. Since the Big Question is not a discipline-specific question (definitely up for debate among many scientific communities), some aspects of relevant disciplines will be explored. My hope is that before the reader reaches the main research in the next five chapters, they will understand why all that research was done and if all those mini-questions were useful in understanding the Big Question.

1.2 Why is the Big Question so hard?

So you're a scientist and you'd like to make a measurement. Great! In order to make a measurement, you'll need to find something you can observe. If you want to measure the temperature of your favorite latte, you'll need to observe heat (or lack thereof) making a change to a thermometer or your hand. So you put your hand on the cup and feel that it is pretty warm. That is one measurement, albeit not very accurate. Trying to be more precise, you stick the thermometer in your cup and it reads 40°C. Another measurement.

The problem with measuring *anything* comes from our ability to make observations. We are currently limited by our bodies and the instruments we make. Without a thermometer, it would be difficult to precisely measure the temperature of a latte and

Figure 2: Once I worked as a busser in a cafe. One of the regular customers asked if I could get him more coffee but he wanted it "Literally boiling hot". So I poured him a fresh cup and microwaved it until it boiled. He was very pleased and drank it right up. Clearly, one's idea of "too hot" varies greatly from person to person (image from captainhowdy27 2016).

we would be back to using our skin as an indicator (not a safe idea). This is one of the biggest barriers in tackling the Big Question, generally because it involves trying to understand biology as a complex system.

It is unclear what part of a complex system to measure because the whole thing is, well, so complex. A good example of this is how American politics changes over time. What main factors account for the election of such-and-such person? Human psychology, economic factors, means of communication, and global affairs are just some of the countless factors that could potentially influence the dynamics of American politics. So if we want to "measure" the likelihood of so-and-so getting elected, what should we observe?

This is actually an old philosophical investigation. Alfred Whitehead studied the relationship between measurements and observation in a series of lectures and

books. He was particularly keen on how this affects our ability to understand living systems. One of our setbacks, he argues, is that we sometimes merely try to invent a model that fits our obeservations (Whitehead 1938 and Whitehead 1934). Merely describing a system, called "positivity", isn't very helpful in describing how biological systems function. Instead, we should try to understand the mechanisms that *cause* our observations to occur.

However, making models is an important first step. Understanding the world in terms of our own experiences may be essential into understanding what life is. In fact, having perspective itself is an important feature in biology. According to Whitehead, the world as described by science means hardly anything in terms of our day-to-day experiences. This isn't to say that scientific models are not useful, but they don't "mean" anything while we decide if its worth ordering Jimmy John's for lunch.

Haldane also chimes in with this quote from Haldane 1947: "But to suppose that one can describe life fully on these lines is to attempt to reduce it to mechanism, which I believe to be impossible. On the other hand, to say that life does not consist of chemical processes is to my mind as futile and untrue as to say that poetry does not consist of words." The term "meaning" is quite loaded because it could imply "purpose", which is unclear and muddled in biology[2].

Aside from that, studying life as a phenomena is self-referential, at least philosophically. Humans who observe biology are also biological, so they are also observing themselves in some sense. Self-reference is a sticky-paradoxical function, such as "This sentence is a lie." In fact, Kurt Gödel showed that self-reference in simple mathematics can lead to only two situations in Gödel 1931: Either a complete set of

[2]Understanding meaning in biological terms is an entirely separate thesis in itself. Thus, it is not small enough to count as a mini-question for the Big Question

mathematical axioms is inconsistent and paradoxical or a set of axoims is consistent but incomplete[3]. Self-reference has also been studied in less abstract systems such as language (Naoto Kataoka and Kunihiko Kaneko 2000 and N. Kataoka and K. Kaneko 2000 are examples). It is actually an important mechanism behind the construction of new words and phrases within an existing dictionary (Levary et al. 2012). But again, it is full of paradoxes and difficulties. Self-reference, along with formal rules of language, allows systems to acquire meaning despite being made of "meaningless" elements. It also can help explain what it means to communicate, how knowledge can be represented and stored, and even what meaning itself means (Hofstadter 1979).

To solve the paradoxes of self-reference, Whitehead steps in again and insists that entities exist on two levels in Whitehead 1928: the physical and the abstract. However, the abstract is derived from its physical components, and this is famously known as the *ontological principle*. Our very descriptions of the physical world are abstract, even if built entirely from physical principles. So can biological abstractions such as function and meaning be captured in terms of science (Bedau 1991)? Why are simple models so successful in the sciences while really complex models (that reflect the real world) are not? Is could be because they have a low effective dimensionality, meaning only a few variables are really relevant for what we are trying to measure. We are able to tweze simple models out from our incredibly complex world simply because of the **scale** from which we make our observations (Transtrum et al. 2015). If we tracked the position and velocity of all the air particles in the room, we could take the average velocity and figure out the temperature. However, given that we don't know all this information about particles, we have to calculate the temperature some other way by making observations on a scale that is much larger than that of tiny particles.

[3]For more reading, see Godel's Incompleteness Theorems, as in Gödel 1931.

Self-reference, paradoxes, scales, meaning, observations, measurements... Somehow these things start to seem related in a mysterious way. Perhaps these are all essential to answering the Big Question because they seem to play a role in biology.

1.3 *Is* life different from not-life?

Suppose it is inappropriate to ask if life is any different than regular forms of matter[4]. Perhaps there is really no distinction between life and not-life except through what we perceive; in other words, "life" could be an artifact of our own perception. But already we have tripped ourselves up! We are exploring this problem with countless preconceived notions affixed in the back of our minds. In fact, it is impossible to remove these preconceived notions simply because we perceive the world in a particular way. Is it possible for humans observers to avoid this self-referential trap?

I've always wondered what it would be like to see the world without human eyes or any sort of sensing thing. What if I were a formless spirit floating about space, like in *The Black Cloud* (Hoyle 1957). Would the color emerald mean anything to me? What about avocados? Does a cat see the sunset the same way I do? Scientists generally remove as many biases as possible by inventing instruments that more than one people can use. Instead of measuring the temperature of a latte with the palm of your hand, it is better to use a thermometer. This removes biases introduced by ambient temperatures and personal heat tolerances (I think 90°F is a perfect temperature

[4]At this point, I think it important to highlight the difference between "life" and "living" for proper context. I refer to "life" as the general phenomenon captured by biology here on Earth, while "living" is the opposite of dead. I've always thought of dead leaves. Obviously not alive themselves, leaves are a clear indication of a nearby tree that was alive enough to make those leaves. Are both life? Yes, the leaves were a part of a living tree which is definitely life. But what about half a leaf? Sure. A single cell of a leaf? Also yes. What about half of the cell? Now how about one protein? An amino acid? It's tricky.

while many would disagree). Thus, when many people measure consistent observations using identical instruments of measurements (like properly calibrated thermometers), we safely feel that we have sidestepped subjectivity.

But things aren't sorted out just yet.

1.4 The number of uses for a screwdriver

I propose the following assumption: Life is a system. One could point to a plug and shout "You're not alive!" and the plug, where notions like dead and alive don't even apply, would just sit there. But the plug *is* a part of life. It was made by humans for other humans and is a part of our living society, much like a termite mound.

Can this idea be applied to things like proteins and RNA? Perhaps proteins and RNA were originally manufactured by life for life, or perhaps they were made by some pre-life chemical system that ended up being life as we know it today. Amino acids have been found on meteors several times (Koga and Naraoka 2017, Glavin et al. 1999, and GLAVIN et al. 2010 are some examples) but can we say that these amino acids are a part of an alien living system? No, though it is a possibility. However, if these same amino acids were found on an ocean planet ruled by jellyfish-people then its much easier to assume that these amino acids are a part of a living system. Note that context is important.

A favorite thought experiment floating around these days is a story about a screwdriver, adapted from a story told by Kauffman in Stuart A. Kauffman 2011 and Stuart A Kauffman 2016. There was once a man whose grandfather passed away. As a part of his inheritance, he received a complete set of sturdy, well-used tools that were manufactured sometime during the 1850's. The man's TV antenna had a poor

signal and he was too poor to afford a new one. Instead of purchasing an antenna, the man fixed an old screwdriver atop of his TV, thus boosting its signal.

Try and count all the possible uses for a screwdriver when it belonged to the grandfather and compare it to the number of uses it has for the grandson. Initially, one might be temped to think the grandson has one more use for it, but actually the number of uses for a screwdriver is totally undefined. In the 1850's, there was no concept of television antennae so the screwdriver could not have been used as an antenna. But as time and technology progressed, the number of uses for the screwdriver changed as well. In addition, technologies that were used in the 1850's were no longer around, which could have decreased the number of uses.

Components of life like proteins, amino acids, and RNA are a lot like this screwdriver. Since proteins have very specific and context-dependent functions in biology, they can confidently be called a part of life. But without the context of the rest of a living system, proteins might not be considered a part of life if they were to sprout up on their own, randomly. What are the number of uses (or functions) for some RNA on Mars? What about amino acids found on asteroids? Of course, these questions couldn't possibly be answered now, but this is meant to emphasize the importance of context.

Defining a universal life could be completely dependent on the context in which a maybe-life is found. Since it's nearly impossible to understand the function of basic objects like screwdrivers without their context, it might be worth pausing that train of thought and trying something else.

The screwdriver story has a second part. "What are the odds that we find a screwdriver on Mars?" (Stuart A Kauffman 2016). What are the odds we find amino acids on meteors? The whole idea of physics is based on the idea that some

particular configurations of the universe (also called states) can transform into other configurations of the universe according to some predictions. These human predictions are our laws of physics invented to match our observations[5]. Not only is context important for understanding amino acids, but also cleverly guessing the odds of them existing. After all, matter has many possible states to choose from, so how did some of it end up as amino acids?

There are several valid philosophies that stand behind probability models, but I think the most relevant and useful one for this dissertation is the computational-based one. Aside from quantum randomness (which the randomness can be calculated deterministically), most of the laws of physics are deterministic and *not* random. The most common image of probability, however, is not much different than someone picking colored marbles out of a bag. If 80 our of 100 marbles are red, then the odds of picking a red marble are 80% because the picker picks marbles randomly. But if the laws of physics are deterministic, then where does randomness come into play when we discuss the odds of finding amino acids? Computation theory takes care of this for us!

Chalmers wrote about the computability of a rock in Chalmers 1996. A rock sitting still deep in the Earth is one particular state the rock can have. If nothing moves the rock, then it will continue to be in the same state. Under Newton's Laws, the rock continuously computes the same state it started in. In other words, given some initial condition (like me winding up to toss a cat into the air with some velocity v at some angle θ), the laws of physics will evolve the state (of the cat) over time

[5]Granted, I'm making a really bold statement here. This is prodding the old debate of whether or not math is invented or discovered (there are hundreds of essays on this topic, such as Abbott 2013). I'm not going to go into it explicit, so I'll simply assert that it is invented for now.

Figure 3: Like rocks, cats can compute trajectories of motion given an initial state and some laws of physics. But unlike rocks, cats shift the shape of their bodies midair so that they land on their feet. The exact physics of cats falling is an interesting research topic (Montgomery 1993).

accordingly (the cat will follow a nice trajectory s through the air as shown in Figure 3 and land on his feet... then run away).

Now the notion of probability has become "how does this state reach this other state" instead of drawing marbles from a bag. Of course, the path from going between certain states could be a random walk if the governing laws behave randomly, but this frames living systems in a more physics-y sense. More importantly, the governing laws of physics can be quantified in terms of what states they can (and cannot) compute, at least in theory. By tossing everything complex about complex systems into state-of-the-universe-A, state-of-the-universe-B, etc, we are able to tweeze out a smaller number of important variables. This implies a loss of information about the specific details of the universe, but it is now much easier to invent bite-sized questions that can help understand the Big Question.

In terms of computers, the laws of physics as we currently know them are somewhat

like software and the initial states are somewhat like data. Imagine a calculator: You feed it an input $80+64$, punch $=$, and the calculator runs the input through a program and you get 144 as an output. But the backend of the calculator is doing something much more intricate in terms of circuits. The circuits must be arranged properly to be able to do math properly. The calculator is in a state that allows it to perform correct mathematics, despite it being made of physical elements[6]. This physical elements could easily be replaced by pegs and wheels and even rocks. With the laws of physics as a computational program, what can possibly be computed here on Earth? It is possible to give a rock all sorts of initial conditions: throw it, lick it, launch it into space, sit on it. Every initial state we give a rock will result in some output. The computability of a rock is practically immeasurable much like our screwdriver, simply because it depends on the system that provides it initial conditions.

Then how did the laws of physics compute life with matter in the universe?

1.5 Emergence

Perhaps one of the most perplexing features of living systems is emergence. Everyone loves to use starlings, shown in Figure 4, as a visual example of emergence because of their uncanny ability to make wonderful patterns as a group (Young et al. 2013). How do they move in such amazing ways? Each bird has its own set of rules, something like "Follow the nearest three birds" and "try to fly parallel to the bird on the right." The rules don't change over time, but do rely on the positions of

[6]This is one of the reasons why I think the laws of physics are invented and not found. Mathematical computations can be performed only in very specific physical configurations, which are encoded *physically* in matter. Perhaps the same should be said about other computations, like the laws of physics.

Figure 4: *Starling Dance*, by Nir Smilga Smilga 2014.

neighboring birds. As a result, the entire group of birds seems to have a motion of its own that emerges from the rules of individual birds. This sort of thing has been reproduced in computer simulations of "boid" agents, originally termed in Reynolds 1987. Figure 5 shows a screenshot of a single timestep of one implementation of a boid model.

However, given the rules of individual agents in a system, there is currently **no way** to predict the emergent, collective behavior. Below are two more detailed examples of emergence in living systems.

There's a really great example of birds migrating south for the winter. During the course of its trip, the bird loses many of its atoms and replaces them with new

Figure 5: *Boids + Perlin Noise Flow Field*, by Nat. Nat 2010

ones. By the time the bird reaches its destination, almost all of its atoms have been replaced. But is the bird a new bird? Yes and no. It is made of new atoms, but the bird is still the same individual (see Mora et al. 2016 as an example of this idea). The emergent macroscopic entity of the bird is unchanged even if its microstate (the individual atoms) has (Vicsek et al. 1995).

The second example is one of a brilliant professor who did amazing research and was loved by everyone, adapted from the essay in Zuboff 1982. The professor was growing very old and it saddened the students and faculty that he was no longer going to be with them. Instead of leaving the professor to natural causes, some grad students came up with a wonderful plan. One afternoon, they drugged the professor so he was knocked out. The took him to a very sophisticated lab and hooked him up to a very intricate machine that covered his entire head and body. There was a

large screen in front of his eyes, so when he woke up, he didn't realize that he had been stuck to a machine and went about his daily life, thinking he was continuing his research and life was normal.

Things were going very well until one of his brain cells died. Carefully, without the professor knowing, the grad students replaced that cell with a tiny machine that performed the same function as the missing brain cell, and things went on happily. This part was maintained by the graduate students, except that more and more cells kept dying. But they were easily replaced by even more tiny machines that were maintained by even more graduate students. About 50 years went by and the professor was happier than ever, so excited that he was the longest living man in the modern era– and so healthy feeling too! Suddenly, one of the millions of grad students who helped maintained the professor noted that the professor was completely replaced by machines! Not a single original cell remained of the professor. The grad student wondered, "Is the professor really alive? If not, at what point did the professor cease to be a living being?"

Understanding the phenomenon of emergence and how macrostates and microstates relate may be one of the greatest barriers in answering the Big Question and other related ideas, especially the mysterious origin of life.

1.6 Information, Computation, Thermodynamics, and Biology

What does a state of the universe look like anyways? There are countless ways of describing the universe at any given time, depending on what is important to measure. At least in the theory of computation, states have been traditionally reduced to abstract sequences of 1's and 0's so theories can be developed easier. Though these

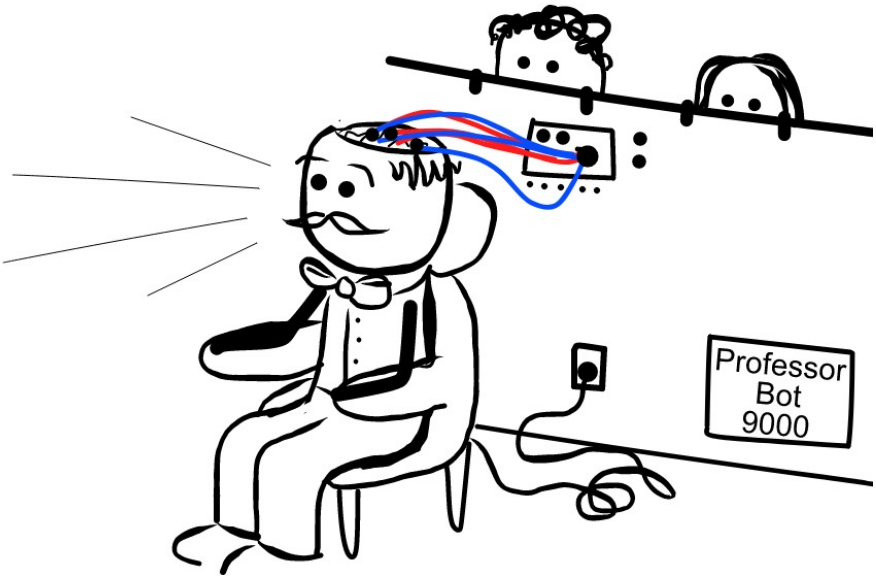


Figure 6: An illustration of Zuboff's professor from Zuboff 1982.

sequences don't quite resemble much in the "real world", they are a very useful analogy of the universe changing according to some law (or computation or running program).

This leads us straight to information theory. The idea of real, measureable information was first invented by Shannon in the 1940's while working at Bell Communication Labs (Shannon 1948). Shannon's formulas for information were particularly useful in understanding how humans send messages through communication channels, like a telegram. By happy coincidence, Shannon's mathematical expression for information was equivalent to the common expression for entropy, with the small difference of a minus sign (Shannon 1948). This got many people wondering how information and thermodynamics could possibly be related (after all, these kinds of similarities hint at some very deep connections).

In fact, computation was thrown into the mix too. In the meantime, Maxwell propsed a thought experiment that baffled scientists for several decades after, better known as Maxwell's Demon (originally presented in a letterKnott 1911). In this

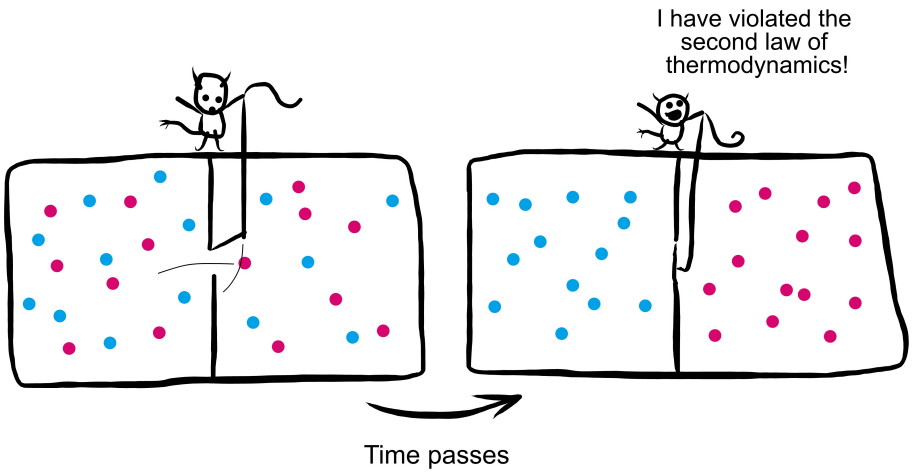


Figure 7: An illustration of Maxwell's Demon.

thought experiment, a sealed box is filled with regular air. The box is separated into two rooms by a wall. On the wall is a little flap. Now, some little demon guy sits on one side of the wall and only lets fast particles into one room, while keeping the slow particles in the other. Pretty soon, one room is full of fast particles and the other is filled with slow particles. One room is hot and the other is cold, creating a pressure from the hot room, thus violating the second law of thermodynamics!

The solution to this problem lies within computation citerex:2017. The demon must measure the velocity of particles headed for the flap and then makes a decision on whether or not to open the flap. After the decision has been made, the demon must erase that bit of information from its memory to make another measurement and decision. After some calculating the minimal energy needed to compute a decision and erase the bit of information, it turns out that the amount of heat produced from the demon's mind is exactly the same as the work gained from having separated hot and cold particles! As a side note, this is the main reason computers produce heat

(Rex 2017). Most of the heat is generated from erasing information from the memory sticks rather than the battery or electrical box.

Although the relationship between computation and thermodynamics is fairly clear, it is still horribly unclear how information and computation has a physical role in the biological world, particularly in terms of heat production and energy. It is certainly possible to calculate the minimum energy required to perform the smallest possible computation (also known as Landauer's limit) by consuming a single bit of information (Landauer 1961). But it is unclear how these simple models apply to real-world biological systems such as DNA transcription and translation. At any rate, an apparent relationship is forming between theories: Information $\rightarrow$ Computation $\rightarrow$ Thermodynamics $\rightarrow$ Biology. So far in this discussion, the second arrow is the most established.

1.6.1 Algorithmic Information

In Zenil 2012, Zenil suggests that algorithmic information is a possible bridge between connecting information, computation, thermodynamics, and biology. In particular, some think that natural selection is actually the extraction of information about an ever-changing environment and encoding it into genes (Vladar and Barton 2011). Although this makes intuitive sense, the actual physical mechanisms (especially in terms of nice equations) of this information exchange are completely unknown. While it may seems a step backwards, algorithmic information abandons Shannon's idea of information. Instead, this method uses a things-are-created-by-programs approach which better connects the relationship between information and computation (the first arrow).

Algorithmic complexity is understood as something like this:

- Short, simple strings of 1's and 0's (like 101010) were most likely made from a small, simple computer programs. These have low algorithmic information and complexity because the program is small.
- Long, complicated strings (like 1001010111010101) were very unlikely made by simple programs and were probably made by longer, more complex ones. These have high algorithmic information.
- A string can only be as complex as the program that created it, never more complex.

So if a string of 1's and 0's is truly totally random, then the size of the programs that created it is absolutely no shorter than that string in terms of the numbers of 1's and 0's the program contains. Probabilities are re-introduced in terms of a marbles-in-a-bag-model. The same string of 0000 can be produced by several different programs, but most of those are short and simple programs (if you are familiar with programming, think of the number of ways you can print 0000). Of course there really very complicated programs that can do the same thing, but they don't occur as frequently. This can be shown by generating every possible program of 1's and 0's and running them on every possible initial conditions of 1's and 0's until they stop ("The Online Algorithmic Complexity Calculator" 2017 has a wonderful interactive online example)[7].

It's really abstract to think of programs and strings of data made of 1's and 0's, but algorithmic information theory can be generalized to anything beyond 1's and 0's. Although this sounds like a fantastic way to measure the complexity of a DNA

[7]Though not possible, this can be done on very large and powerful computers up to programs and states of a certain size to get a general sense of things.

sequence and other biological wonders, the downside is that algorithmic information can actually not be exactly measured. It can only be approximated (Lui et al. 2015) due to that pesky halting problem posed most famously by Turing in Turing 1937. The halting problem basically states that it is impossible to know *how long* a computer program will run until it has stopped (halted). It may be possible to figure out what that program will output before you run it, but it is currently impossible to know how long it takes until after the program has finished running. So basically, for a real biological system, the algorithmic information cannot be calculated.

Although the complexity/randomness of a thing might not be possible to capture even approximately, Zenil and collaborators have already started to invent ways to describe how the dynamics of a system can change over time (Zenil et al. 2017). This proposed calculus suggests ways of intervening on a system to change its algorithmic information. With this, it could be possible to understand (and eventually control) a system's dynamics by increasing or decreasing its algorithmic information content. Perhaps with more development, dynamical systems including biological ones can be understood in terms of computation.

1.6.2 Why Information?

What's with the long discussion on information anyways? Biology is anything but simple to us, yet science thrives on the idea of simple, toy models. Is it possible to make simple models in biology? Of course, we already have excellent models of predator-prey dynamics and insights from agent-based models (as in Fraczkowski and Olsen 2014), and also mathematical models like the Lotka–Volterra equations (Lotka 1910 and Goel 1971), but there is a clear lack of unification of all biological processes.

Is all of biology simply "just messy" and disorganized, or is there a small handful of simply biological laws that can be applied to every aspect of biology? Researchers from many different disciplines have found evidence that information is the key to understanding biology in an elegent, simple way (Davies and Walker 2016 and Walker, Davies, and Ellis 2017).

Biological randomness is not only an essential component of the intrinsic unpredictability of life, due to the interaction between many levels of organization, but also as a key component of stability. In fact, increasing organization invites growing disorder and increases variability and differentiation (Buiatti and Longo 2013). But how is randomness intrinsic to the system in question? I mean, assuming that all of physics is governed by deterministic processes, perhaps randomness is simply lack of a complete picture in the biological entity. Someone picking colored marbles out of a bag experiences randomness from their perspective, but there is nothing random happening from a larger, systemic level. The picker's hand choosing a marble obeys non-random laws of motion. Clearly randomness has a large affect on biological processes but it is unclear how.

This is where information comes in. Biology is full of function and semantics, which sounds like the perfect place for information theory. The role of information in living systems should be taken seriously, especially when it comes to the origin of life. Perhaps the transition from non-life to living systems was a direct result of matter being able to utilize information (Walker and Davies 2013) in a way that has a meaningful effect on that matter. In fact, information heirarchy deynamics are currently replacing the model of evolution as a passive filter selecting for random changes on the genetic level (Walker et al. 2013). This will help us understand how

cells and organisms work (Nurse 2008), and likely many other levels of biological organization.

John Wheeler also shares similar sentiments, but more specifically in terms of quantum theory and information (Wheeler 1990). He discusses the idea "it from bit," which means that every physical quantity derives its ultimate significance from bits of information, including its actual physical existence. Here, Wheeler uses the original definition of bits of information: yes or no, 1 or 0. In other words, all of reality is comprised of yes-no responses and this is the information that defines our reality. This is another way of understanding how information might be instantiated physically, as with biology.

A controversal topic in the information discussion is the role of maximizing entropy in physical systems like biology. According to the principle of maximum entropy, the most likely probability distribution of several evens happening corresponds to the distribution with the maximum entropy, or most randomness/unknowability associated with it. This notion is used in artificial intelligence models by means of inference engines, which apply logical rules to the knowledge base to deduced new knowledge Hayes-Roth, Waterman, and Lenat 1983. This is useful for gaining inference from complete or incomplete information (Jaynes 2003). This perspective may be interesting in understanding the mechanisms of biology, since each layer in a multi-layered system processes incomplete information about the system as a whole. Jaynes thought that probability theory was an extension of logic and has real physical implications, particularly in thermodynamics. Thus, the physical thermodynamics of a biological system can be understood in what each piece "knows" about, or how much information about the whole system it contains.

Information is not a thing that happens in a single lone element of something

floating along in space. Traditionally there are message senders and receivers and a medium for the message to travel through. Both the sender and receiver filter and distort the message in some way, which is further distorted through a medium. Causality can be useful in understanding how information flows through a system of senders and recievers. Recent results suggest that epigenetic (external) information cannot be reduced to genetic (internal) information in biological systems. For example, in individual organisms, development is the expression of information accumulated during evolution and heredity is the transmission of this information (Griffiths 2017). Information involves very long time scales and several intricate factors.

Many argue that causes differ in the degree to which they are "specific" to their effects. Causal specificity (it's called) can be measured but not without a probability distribution over the states of whatever variable is expected to be the cause (Griffiths et al. 2015). But it is also useful to have a theory of information expressed solely in terms of which transformations of physical systems are possible and which are impossible. Constuctor Theory (discussed later) of information regards information as something whose nature and properties are determined by the laws of physics alone (Deutsch and Marletto 2014), simply by understanding how one state of the universe can transform into another. This is also heavily related to understanding what causes what in physical systems, especially since the laws of physics must be expressed to be consistent with all these notions of information.

Even though having a notion of information is important, meaning usually isn't included in the information-theoretic models of biological communication. Some theories of biological information use functional responses as a replacement for probabilistic descriptions of correlations between sent and received messages. This poses a problem because it leads to potential paradoxes, such as having the most amount of

information associated with a channel that creates completely wrong interpretations of messages. To remedy this, inspirations were taken from the concept of duality of the communicative sign stated by the swiss linguist Ferdinand de Saussure. This allows for a complete description of the minimal system necessary to measure the amount of information that is consistently decoded (Corominas-Murtra, Fortuny, and Solé 2014).

Within the quantum mechanics framework, things that leave behind a "trail of information" do so by increasing or maintaining the same amount of entropy. Events where the entropy actually decreased haven't left behind any information about that event, which is actually completely indistinguishable not having happened at all. Therefore, physics cannot study those processes where entropy has decreased (Maccone 2009), having some very real biological implications.

Of course, the efforts of understanding biology in terms of information has not gone without criticism. In particular, a main criticism is that it encourages genetic determinism and the application of these theories hinders the understanding of organisms in the lab (Longo et al. 2012). This is a fair assessment since many current information-and-biologically-relevant mathematical models do not have much direct application in the lab. However, there have been several results from the lab suggesting information may indeed play a vital role and that it should be understood clearly.

1.6.3 Results from the Lab

Information processing in biological organisms can probably be done with different qualities, ranging from these-mice-can't-see-cats to these-mice-can-geo-track-all-predators, but how can we measure it? Also, how do organisms improve their information-processing throughout evolutionary processes or not (Nemenman 2012)?

These are just some of the questions that biologists (and many others) are asking pertaining to only one biological level of organization. The trouble with biology is that there are many many layers of organization and interactions. It is difficult to make generalized statements about biology as a whole without understanding the smaller layers first.

On the level of an organism, organisms may have complex structures because they need to be able to process information about their complex environments. As all organisms in a biological system grow more complex, the environment, according to a single organism, also becomes more complex because the environment also contains complex organisms. Lots of complexity nested withing complexity! But what are the energy costs in maintaining these structures capable of processing information, such as brains or eyes? Surely this should relate to the energy costs of computation. Some models suggest that the ability to accurately predict the environment outweighs this energy cost (as in Seoane and Solé 2017) so processing information might actually be energetically efficient in the long run.

In addition to these energy costs, organisms that are capable of regenerating their body parts must be able to store blueprints and executing instructions in some way. Planaria flatworms have been extensively studied for these reasons at the Levin lab (a comprehensive paper on this is Lobo, Beane, and Levin 2012). By studying the interactions between proteins that are present throughout the worm, regenerative networks have been discovered as the "information storage centers" and "programs" that allow the worms to regenerate their lost parts (Lobo and Levin 2015).

One way to test the role of information in living systems is to compare some aspect of an organism to something that is random. When the gene regulatory network of fission yeast is compared to a random network of the same size (with the same

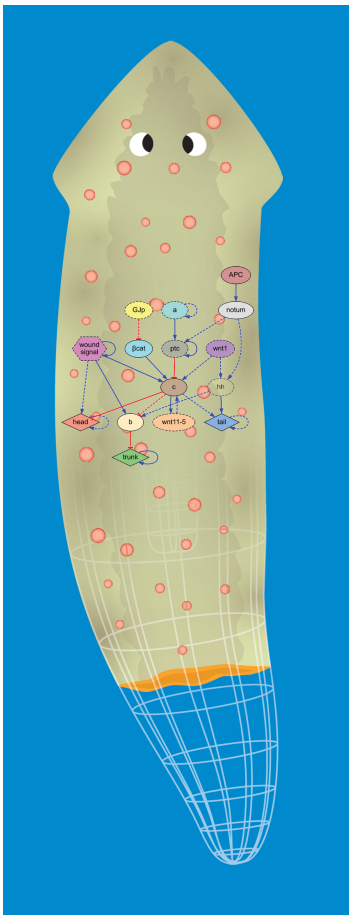


Figure 8: Planaria flatworms are able to regenerate parts of their body that have been lobbed off, including their head. How does the worm know how to do this? Somehow, information about its entire structure is stored throughout its body. A network of interacting proteins has been discovered via machine learning as a possible mechanism for regeneration (Lobo and Levin 2015).

number of conenctions), the biological network processed and store information while the random networks did not (Sara Imari Walker 2015).

On a cellular level, studying information also leads to insights on cancer. Some suggest cancer is the result of a system disorder of the algorithms that governs the regulation of living cells (Levin 2012). These algorithms would normally control cell activities in a way that is specific to their particular organs (Moore, Walker, and Levin 2017). But cancerous cells could have defective programming that results in selfish behavior. Zenil's calculus could be applied to this system to understand how algorithmic information can affect the behavior of certain cells (Zenil et al. 2017).

On a social level, such as baboon societies, language could have evolved from the social knowledge of baboon relationships (Seyfarth, Cheney, and Bergman 2005). These relationships are heirarchially structured. Typically in biological systems, heirarchical structure imples some global transfer of information of who is more dominant than another. Studies on zebra finches also have shown similar results (see Ma, Maat, and Gahr 2017).

On an evolutionary level, Fisher information becomes the intrinsic metric of natural selection and evolutionary dynamics. By maximizing the amount of Fisher information about the environment, a population leads to Fisher's fundamental theorem of natural selection, as seen in Frank 2009. Regarding environments, making note of important features and ignoring other features comes at a cost to an organism. If the organism does not notice particular things, it could cost the organism energy when solving difficult problems. However, maintaining sensory organs also comes at an energetic cost. Which costs more? When errors are allowed in environmental senseing, very efficient internal representations can be discovered by selecting the next most efficient thing, one generation at a time. When the complexity of an environment increases,

organisms that use efficient algorithms are able to detect the subtle distinctions in the environment (Marzen and DeDeo 2017). Maximizing the amount of Fisher information may be the most energetically efficient solution to this.

On a thermodynamic level, biology is a chemical and physical process. It is important to understand it in traditional thermodynamical terms like energy costs. As such, there is a need for some theory that can capture chemical information processing in terms of energy constraints (Smith 2008). By using Landaur's principle, reaching the limits of the energy costs of chemical information flow is entirely possible, even if the literal piece-by-piece assembly of the same system is not energetically feasible. Real chemical reversible models of energy and information flow can be easily achieved within ideal energy limits and are related to the fundamental operations of computation simply by understanding the energy costs of information processing.

Finally, on a theoretical level, "consciousness" is hypothesized to be a measure of how casually connected a network of elements is (see Tononi et al. 2016). This can also be understood with the phrase "the whole is greater than the sum of the parts". The information processing of the entire network has a larger "consciousness" measure if that of the sum of its parts has a smaller measure (Oizumi, Albantakis, and Tononi 2014). Of course, the interpretation of this measure and whether or not it can actually be called "consciousness" is a hotly debated topic. Regardless, these methods have brought some interesting insights into how biology processes information.

1.6.4 Machine Learning

Machine learning is also making contributions to the way we understand information processing and biology. Most machine learning systems are biologically inspired,

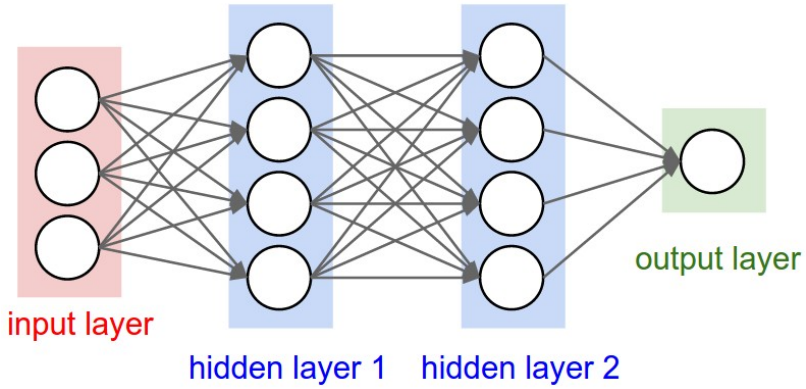


Figure 9: A simple cartoon representation of a neural network (image from Karpathy 2017). The input consists of some configuration (like 1's and 0's) in the input layer, and this affects the states of the nodes in the hidden layers. The rules that determine the states of the nodes besides the input nodes are determined by the states of the previous nodes and the edges. Ultimately, the output layer is in some final state.

particularly neural networks like the one shown in Figure **nn** These networks were designed to mimic the network of neurons in a brain (although nobody really knows the full details of that either).

When the neural network receives some input from one side of the network, it is fed through a series of "black-box-y" hidden layers to create an output. Usually the neral network is terrible at outputting good answers at first, but it can be trained by modifying its connections (for an in-depth overview of these methods, see books such as Du and Swamy 2013). There is a huge variety of training techniques, but the general idea is that the network is changed so that given an input, it gives an answer that is closer to the desired output. Neural networks have been trained for a wide variety of useful and interesting tasks, including mimicking real protein interactions (Ling, Samarasinghe, and Kulasiri 2013).

Information theory has been making a more popular appearence in machine learning, though it is not a new idea in the field. In 1999, Naftali Tishby and

colleagues theorized that general learning occurs during an information "bottleneck" across the neural network (Tishby, Pereira, and Bialek 1999). This is the same concept as information loss, forgetting several irrelevant details and preserving only the important ones. This is the entire basis of thermodynamics, commonly refered to as "coarse-graining", meaning the details are blurred into a smoother picture. However, the actual notion of information processing wasn't something that could be measured in a quantitative way as a neural network is trained. But as of 2017, models finally concretely support the idea of an actual information-bottleneck in deep-learning machine learning (Shwartz-Ziv and Tishby 2017). This is an exciting development for people outside of machine learning because it presents a concrete example of information playing a critical role of an evolutionary (albeit artificial evolution) process.

1.6.5 Constructor Theory

A discussion of information would not be complete without some words on constructor theory. Constructor theory is a new approach to formulating the fundamental laws of physics so that they are general enough to include compuation and information. Instead of describing the world in terms of trajectories, initial conditions, and dynamical laws, constructor theory presents laws that include possible and impossible physical transformations between states (Deutsch 2013). The laws of thermodynamics can be recast in terms of a non-approximative, scale-independent distinction between work and heat. This implies a relationship between information theory and the first law of thermodynamics (Marletto 2016).

Constructor theory may be useful for understanding biological systems. Gene

replication and natural selection can only occur (without the design of biological evolution being directly encoded in the laws of physics) if the laws of physics have certain other properties. The appearence of design in a no-design universe and the logic of self-replication can be described under construction theory (Marletto 2015). These things can exist only if these laws allow information to be physically instantiated. In addition, an accurate replicator must also be a "vehicle" that also acts as a self-reproducer.

In quantum theory, particularly unitary quantum theory which is non-probabilitstic, the unpredictability of measurement outcomes yields superinformation theories. This is also true of the appearence in randomness in a finite amount of measurements. Constructor theory corrects misconceptions about the standard methodology of scientific testing that is inapplicable to Everettian quantum theory, hence the theory is untestable. This is from misconceptions about probability and the logic of experiments. Constructor theory eliminates everything probabilistic from fundamental physics (stochastic processes) and from the methodology of testing (Deutsch 2016).

1.7 More on Computation

Most of us (probably) think of computation in terms of computers. The origin of computers as we know them today is very closely related to the laws of physics in terms of states and transformations between them. So, what *is* computation, beyond the realm of computers? Is it possible to think of computation in terms of the laws of physics? Actually, this is an extremely useful perspective for understanding how our universe changes over time, outlined in Zenil 2013. But more importantly, it is useful to think about *how* nature computes. Does nature use a computer program?

How does it read that program and how does it read data into the program? In any case, it goes to show that computation is an extension of information in the sense that computation is the mechanism that *utilizes* information.

Computers are ultimately physical objects, made of electronic components and circuits. We know them so well that we are able to reprogram computers in several different ways. With the same computer, we can run Mac OSX, Windows, or Linux and the same video game on all three (except if you have Mac OSX). We can use the same electronic components for several purposes, like making radios or perhaps some sneaky device that can actually turn down your neighbor's super-loud 3am speakers (see Gordon 2012) (disclaimer, that is illegal by the FCC, I checked!). It's even possible to forego the digital parts altogether and make simple computers out of cardboard boxes. Charles Babbage even began designing these "engines" in the 1800's (Museum 2017) (shown in Figure 10) and their original conception was discussed in the 1780's in Müller 1786. In fact, the very first computers were made of gears, cogs, and pegs and could be powered by turning a handle, like in Figure 11. So theoretically, why couldn't we re-program any and all bits of matter around us? Re-programming is even more powerful than blunt transformations because computers are capable of making those desired transformations on their own, without us needing to do more than powering them and giving them a starting point.

Zenil argues in Zenil 2017 that we have already been doing this. 3D printers are a nice example of how we reprogram plastic into little sculptures or paperweights. If we are able to reprogram plastic, metal, and wood, then how long will it take until this is extended to biological systems? We have already started doing this with CRISPR gene editing (see Cong et al. 2013). Although one could argue that we've

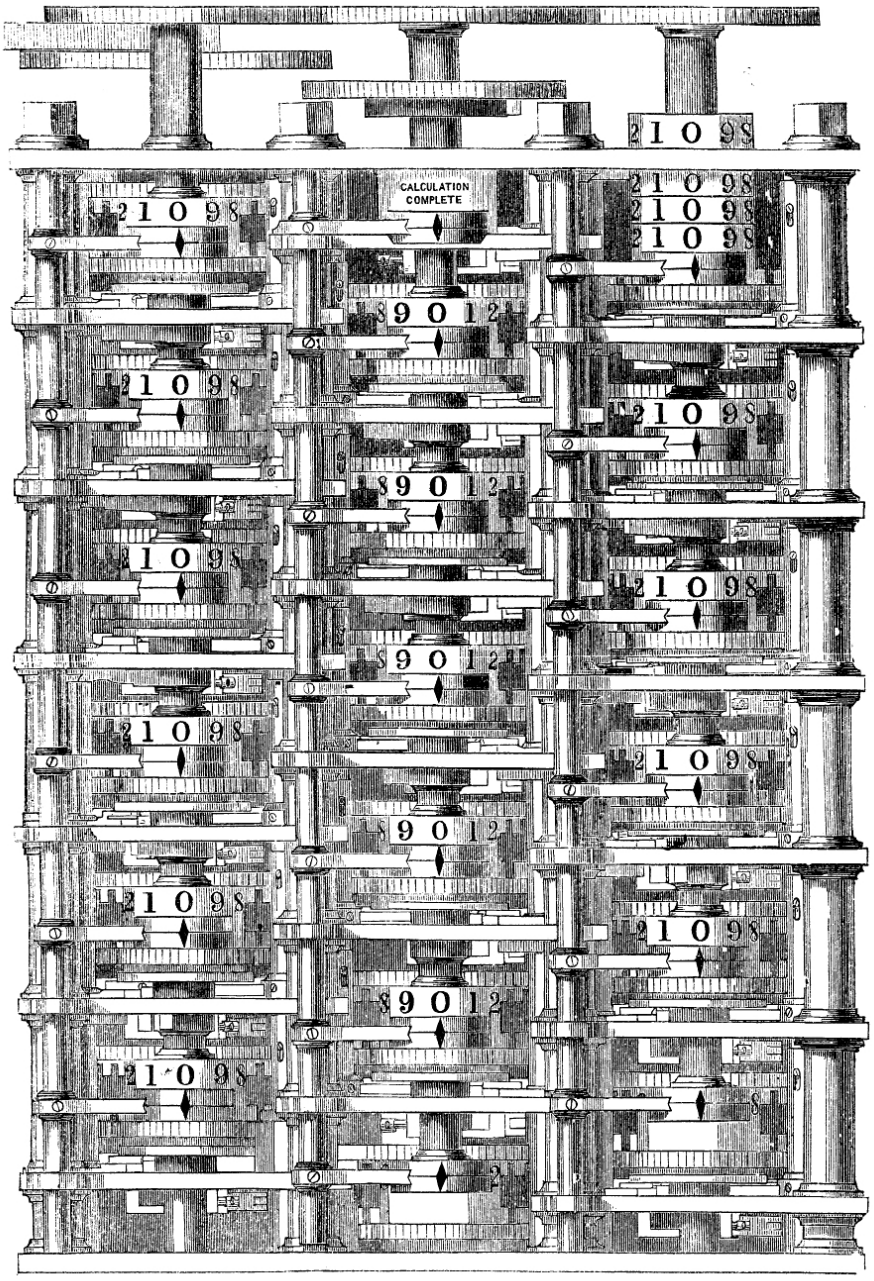


Figure 10: *Portion of Babbage's difference engine*, drawing by Charles Babbage in 1864 from Babbage 1864. Although Babbage's different engine was designed in the 1800's, it was never built until 1991 when it was finished in celebration for Babbage's 200^{th} birthday. The difference engine inspired Ada Lovelace to develop modern theory of computation, which in turn inspired Alan Turing.

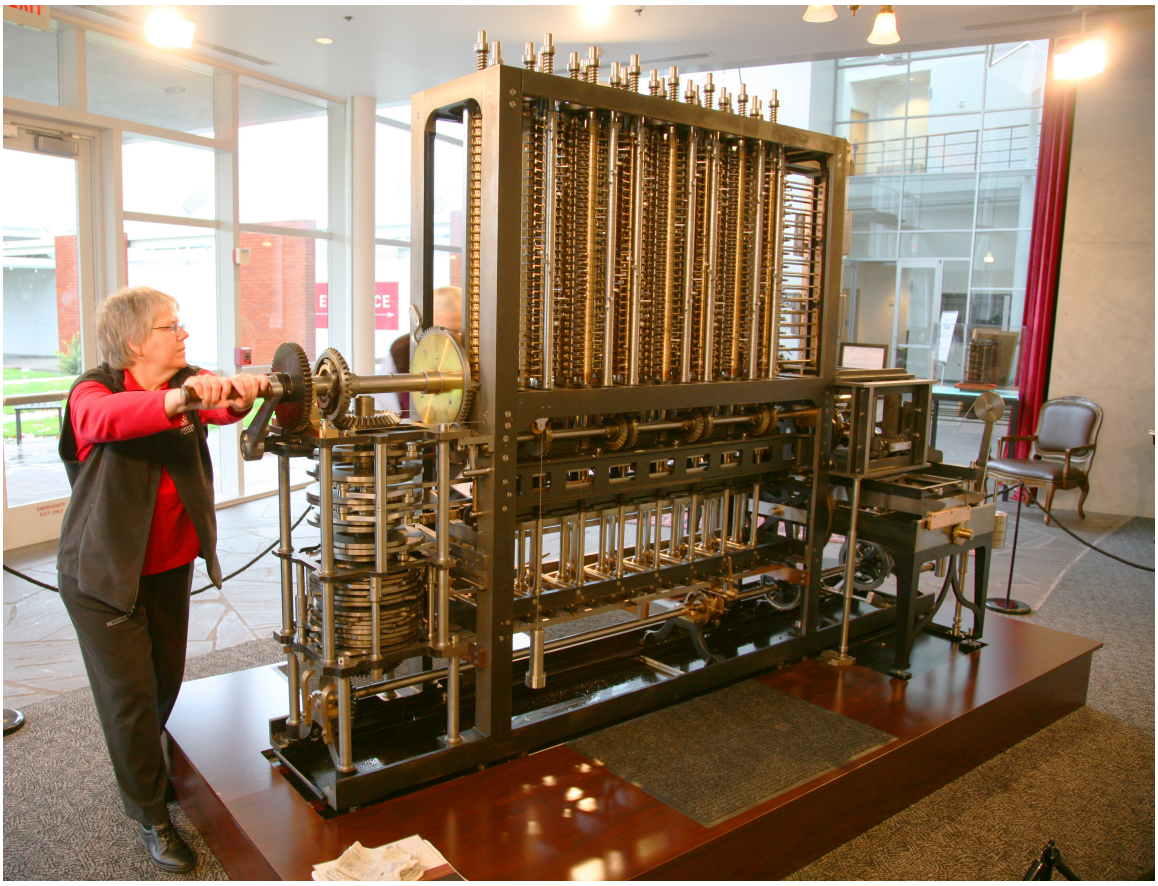


Figure 11: The difference engine at the Computer History Museum is fully operational (image from Briggs 2016). It can be used by punching some bottoms for input and turning a lever until the answer is computed.

been trying to reprogram our own bodies for thousands of years before, simply by trying to live longer and healthier. With the dawn of vaccines, we've been modifying our bodies' immune systems to overcome life-threatening diseases. But to ensure the future success of reprogramming living systems such as ourselves, in hopes to cure cancer and future epidemics, it is likely that understanding reprogramming matter in a precise, mathematically theorized way will help.

To approach the idea of computation in terms of biology, it is important to think about the following things (Mitchell 2011):

- How is information represented in the system?
- How is information read and written by the system?
- How is it processed?
- How does this information acquire function (or "purpose," or "meaning")?

None of these ideas are very straightforwards because the idea of computation has mostly been implemented in terms of computers. Yet in some way, cells and collections of cells can be thought of in terms of little computers. Specifically, bacteria can be seen as computers making computers, in reference to a Turing-complete system (Danchin 2009), meaning they are programs capable of replicating themselves over and over. This discussion isn't actually that far from the discussion of information, particularly algorithmic complexity from earlier.

1.8 Levels of Biology

Biological systems appear to operate on multiple levels of organization, both physically and informationally (Walker, Cisneros, and Davies 2012). Traditionally, we have often thought of the physical world from a bottom-up approach: things happening on the smaller scale compose things happening on the larger scale. While there is nothing wrong with this, there may actually be a bigger picture, especially when it comes to biological systems. One such approach is the top-down approach, where the higher levels of organization actually affect the lower levels. Both bottom-up and top-down flows of information seem to be important in understanding the dynamics of biology (Sara Imari Walker 2014). Because life appears to be different than non-life, a completely bottom-up approach to solving the mystery of life is inadequate (Sara I

Walker 2015). Life may be more than just complex chemistry and physics interacting in more and more complicated ways; it might be something totally different.

In may cases, the coarse-grained view is more "useful" than the microscopic view. In other words, a good macro-view of a system contains all the relevant information and very little or none of the irrelevant information. This way, the higher-levels contain only useful information (Hoel 2017). But in this sense, "usefulness" is arbitrary and depends on the observer.

Models of the heart indicate that top-down causation can be represented as the influences of initial and boundary conditions differential equation solutions used to represent the lower level processes (Noble 2012). From this, we can assume there is no "best" level of causation to describe biology. In fact, there are many levels of causation, and all are necessary to account for complex biological phenomenon.

Five different classes of top-down influence have been identified from real-world observations by Ellis. They are: algorithmic top-down causation; top-down causation via non-adaptive information control, top-down causation via adaptive selection, top-down causation via adaptive information control and intelligent top-down causation (the effect of the human mind on the physical world). An ongoing battle between bottom-up and top-down approaches has important consequences for medical and healthcare issues, as discussed in Ellis 2011.

Another example of where coarse-graining and top-down causation is necessary is understanding how the power grid operates. Understanding the power grid as a complex system requires several layers of how information is processed. On a base level, there is the physical system itself, made of wires and transistors. Communication is a second layer where agents act to maximize their own sources of power by receiving information from their neighbors. Lastly, decisions as a whole system are made to find

a stable coniguration (Kühnlenz and Nardelli 2016). The effectiveness of describing these layers can be tested in an agent-based model, where the behavior of the system as a whole is influenced by the size of the system.

1.9 Dissertation Outline

Where has all this discussion lead to? The entire point is to expose the motivating ideas behind the research in the subsequent chapters. Without this prior discussion, the research seems entirely unconnected to the Big Question at first glance. Video games, universality, cellular automata, and repeating patterns are all bite-sized research topics that are specifically engineered to guide the literature towards the Big Question. In particular, Chapters 3, 4, and 6 are adapted from original research publications (A. Adams et al. 2017, A. M. Adams et al. 2017, and Adams and Walker 2017 respectively). I've added two additional chapters (Chapters 2 and 5) for more discussion that is specific to their following research chapters. Then, of course, this dissertation is concluded with a final chapter of speculative, wild-idea discussion (as if that hasn't been covered already).

1.10 All Models are Wrong

Finally, it's important to reiterate the famous quote "all models are wrong, but some are useful" from Box 1976. All math, physics, and notions learned from science are models we use to describe the world that we observe. However, there's a very old and deep argument that asks whether math is invented or discovered. One way of putting this is the following: Imagine humans are discovered by an alien race.

Would they agree with our mathematics or simply view our math as a product of our culture? I take the following stance: Models are tools invented by observers. "Laws" can certainly change over time and may very well be different in other parts of the universe. But in biology, laws are not static and change over time. Biology is constantly rewriting its own rules, thus making it very difficult to write equations and things for in the sense that we are familiar with. So, while reading these chapters, keep in mind that these models are very limited in the sense that a single program (algorithm) computes them. Their laws of physics and rules, ultimately, are static. In this sense, they may all be fundamentally wrong.

Chapter 2

PRIMER ON CELLULAR AUTOMATA

When I was about 7 or 8 years old, my parents bought a brand new Gateway computer that ran Windows 95. It was glorious and included all kinds of exciting new games. I enjoyed each of these games except for one: The Game of Life. It was a horrible game. Basically it was a giant grid of colored squares that changed over time and looked absolutely nothing like life. Figure 12 is an example of what I saw, but with only two colors. There was no winning or losing or any indication of what was even happening. I looked at the rules that were provided, but those just made the game even more confusing because they were written in some coded computer language. So I decided that a player lost when all squares turned the same color and eventually forgot about this "awful game".

It wasn't until graduate school that I encountered The Game of Life again. It turns out that this was not actually a game, but a scientific model for living systems invented by John Conway (popularized in Gardner 1970). Quite an interesting thing to include in a default operating system in 1995, but I digress[8].

The Game of Life is a two-dimensional cellular automata (CA) model with a very specific set of rules. CA were originally invented as a simple computer model that mimics biological evolution through some rules, described in detail later. They were first introduced by Ulam and Von Neumann in the 1940's to understand how biological evolution can produce regular or irregular behavior from simple rules (for a

[8]This could have been due to the rather large cult following it attracted since its invention in the 1970's, see Rendell 2000

Figure 12: One state in an implementation of Conway's Game of Life (image taken from Life 2017). Cells are either black or white (or 1 or 0, respectively) on a two-dimensional grid. The color of each cell in the next time step will depend on the current state of the cell and all its neighbors.

comprehensive history on cellular automata, see Schiff 2011). In fact, CA have also been known to exhibit kinds of self-reproduction as in C. G. Langton 1984.

In the original Game of Life, a state consists of a two-dimensional grid of black and white cells. A cell changes its color (or not) at the *next* time step depending on the *current* colors of its neighbors. This is the basic idea behind a local update rule: The color of every cell in a grid is determined by the current colors of other cells. The Game of Life is a specific set of these rules, which results in all sorts of interesting little patterns to emerge. There have been numerous studies on these patterns, ranging from "glider" interactions and creations, to giant, stable oscillating structures.

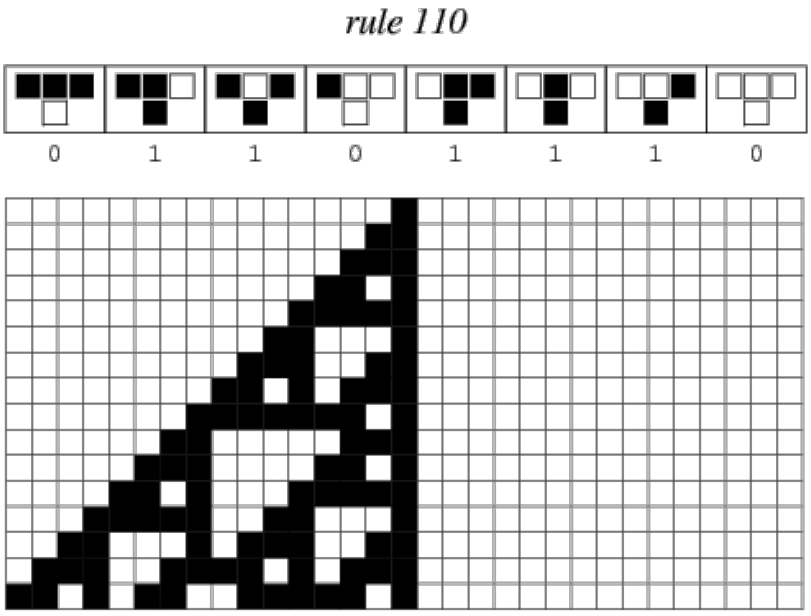


Figure 13: ECA rule 110 and its evolution from a single black cell in a space of white cells (image Weisstein 2017).

2.1 Elementary Cellular Automata

Since cellular automata are very closely related the the ideas on computation in the previous chapter, particularly rules, programs, states, and outputs, they have become a very popular mode. Stephan Wolfram extensively studied CA on a one-dimensional grid in his book *A New Kind of Science*, Wolfram 2002. One-dimensional CA have been dubbed elementary cellular automata (ACE) because they are the simplest symmetrical iteration of CA. ECA only use two colors, black and white, or 1 and 0 respectively. The color of each cell only depends on its own current state and the color of its two closest neighbors. One example of an ECA rule and its implementation on an initial state is shown in Figure 13.

In ECA, the rules are labeled by a number which represents its enumeration of all the possible rules. Since there are two different colors and three cells in a

neighborhood (the one cell and both of its neighboring cells), there are $2^3 = 8$ neighborhood possibilities. Since there are two possible outcomes for the middle cell in the neighborhood, that means there are $2^8 = 256$ possible ECA local update rules. In Figure 13, Rule 110 is represented by one of the 256 possible outcomes for the outcome cells of the neighborhoods. Since these are 01101110, it is named Rule 110 because the binary translation of 01101110 is $0^7 + 1^6 + 1^5 + 0^4 + 1^3 + 1^2 + 1^1 + 0^0 = 110$. Likewise, if all these outcome cells were 00000000, then it would be named Rule 0.

Depending on the initial configuration of 1's and 0's and the local rule, ECA could evolve into many types of patterns. In Figure 13, the initial one-dimensional state consists of all 0's and a single, middle 1. This is the row of cells at the top of the image, with time going down. Many possible patterns were classified into four distinct classes by Wolfram. Class 1 ECA rules are very simple and stay in a fixed attractor state, meaning they find one configuration and stay there forever no matter the initial state. Class 2 get stuck in simple repeating patterns for all initial states. Class 3 look much more random, such as the famous Rule 30 (shown in Figure 14 with the same initial state as from Figure 13). Finally, Class 4 are known to be complex, somewhere between random and structured, depending on the initial state. Rule 110 is known to be Class 4 and by analyzing many different initial states and resulting patterns, it was found to be computationally universal. This means it is able to simulate any other ECA rule given an infinite amount of time and an infinitely large initial condition, proven in Cook 2004.

However, this classification scheme wasn't rigorously defined and was introduced more as general guidelines. Since then, many have tried to find a more precise classification scheme for the ECA rules. One scheme uses an approximation of Kolmogorov complexity, outlined in Zenil and Villarreal-Zapata 2013. Another is a

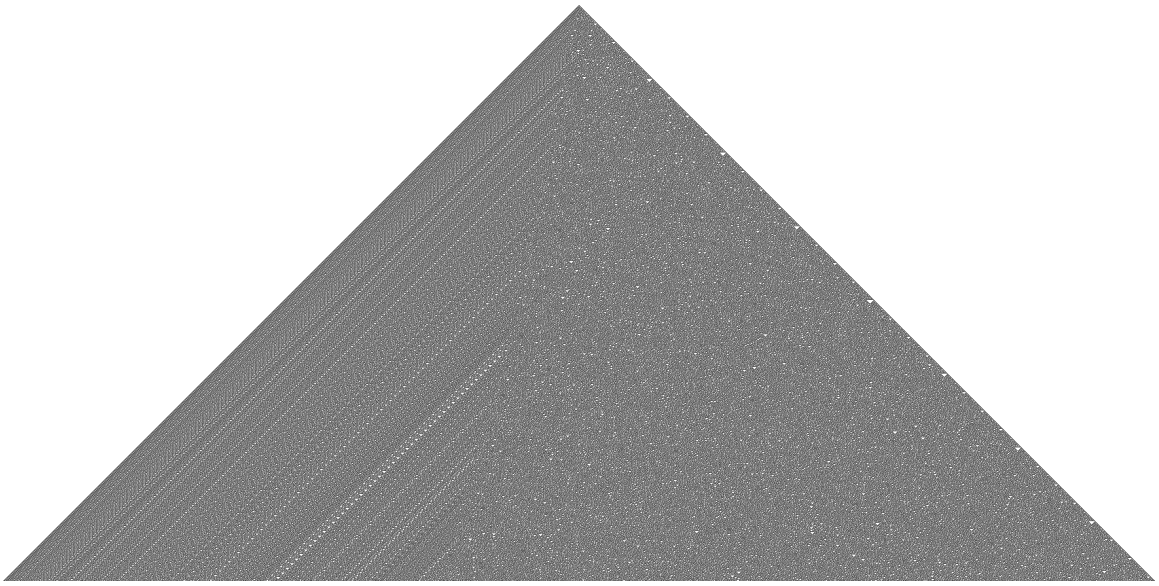


Figure 14: 2000 time steps of Rule 30 from a single black cell in a space of white cells (image from "PiAndWhippedCream" 2008).

compression-based investigation based on the sensitivity of the CA to initial states (Zenil 2010). Applying transfer entropy (a measure of information processing) is also a useful scheme for ECA (Borriello and Walker 2017) because it classifies rules based on their ability to process and store information. These classification schemes allow us to understand what each ECA rule is able to compute.

Besides classifying ECA rules, CA are studied for their original purpose: To understand biological systems. This is an excellent implementation of some of the ideas discussed in the previous chapter if used as an analogy for the emergence of complex behavior (CA states and patterns) from simple interaction rules (laws of physics). If the rules are well-understood, it is possible to measure the resulting behavior of the CA.

For example, it is possible to measure CA sensitivity to changes in the initial state, as measured by the Lyapunov exponent. The Lyapunov exponent can be approximated by flipping the color of a single cell in the CA's initial condition and then measuring

how many cells have been altered from the original pattern as a consequence. Of course, this is partially dependent on the exact cell that is changed, so all possible initial cells should be changed and then the average outcome should be calculated instead (Tisseur 2000).

2.1.1 A Few Variants

Cellular automata have been used in genetic algorithms to find specific, more complicated rules that can solve particular tasks, known as the EvCA project (Group 2000 and Hordijk 2013). A genetic algorithm is a computational process inspired by biological evolution as we currently understand it. Basically, using CA as an example, a pool of random CA rules and initial conditions are generated and executed on several initial states. The algorithm looks at the outputs and picks, say, the best 10% according to some specific criteria, like the 10% most black outputs. Then it splices and swaps those best rules much like in swapping genes in reproduction. This process is repeated several times until a population of rules that is very good at solving the specific task is reached.

Genetic algorithms are particularly good for CA that have so many rules that it is impossible to test them all. In general, genetic algorithms are good for finding a solution in *any* space that is to large to explore completely. Using a genetic algorithm can help find a good solution to a problem without exploring every single possibility. Although they can return good solutions, they certainly do not guaranetee optimal solutions (Mitchell, Crutchfield, and Hraber 1994) but they are general enough to solve a wide variety of problems (Lovinger et al. 2014 and Breukelaar and Bäck 2005).

The EvCA project was a major milestone in understanding how biology can select

for advanced computational tasks by means of using simple local interaction rules (Hordijk 2013), like those of the laws of physics. One of the tasks that was used is called the density classification task. For an initial state of black and white cells, the density classification task wants the systems to turn white if the initial condition was mostly white, or black if the initial state was mostly black, as in Figure 15. In one EvCA model, the bulk of computation was performed by interacting "particles" in subsequent CA states. These particles are boundaries between uniform patterns on a CA grid. The interactions of the boundaries facilitated the exchange of information (Das, Mitchell, and Crutchfield 1994) and the density classification task was solved effectively as a result. These results have demonstrated an example of information transfer when a system is evolved to perform some sort of computation.

Genetic algorithms are also popular for creating digital art (see Figure 16 and Draves 2017 as an example) and music. In fact, the internet radio website Pandora uses a genetic algorithm to pick songs based on a user's preferences (see Inc. 2017 for a complete description). With every interaction from a user, the process that selects the next song is generated by a new generation from the genetic algorithm.

There are several, practically countless CA variations that have been invented to test some aspect of dynamical systems, whether its biology, geology, or political voting. One final example in this discussion is HetCA. This CA variation allows evolutionary progress to be measured with three criteria: the robustness, size, and density of generated genotypes. The results show that the oldest genotypes in terms of evolutionary time are frequently the most robust, and that newer genotypes correspond to a much stronger phenotypic density (Medernach et al. 2015). In general, results

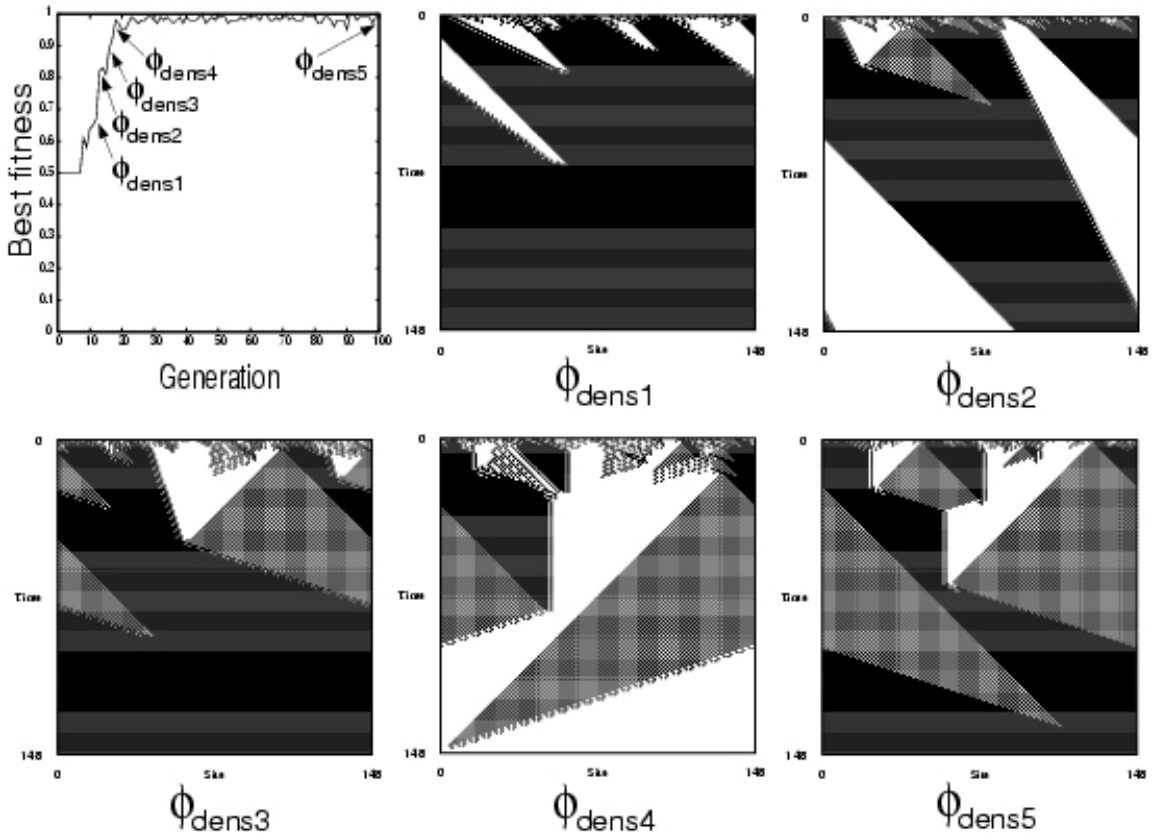


Figure 15: Some results from the EvCA project. These panels are from a single genetic algorithm run. The arrows in the upper left panel indicate five different generations in the genetic algorithm, and examples of these generations are shown in the subsequent panels.These five panels are space-time diagrams with time going down (image from Group 2000)

from CA models are consistent with our observations of the real world and are excellent tools for crafting analogies with biology.

Chapters 3 and 4 are centered around another version of CA that is tailored to understand open-ended evolution. Using CA makes some questions answerable in terms of a limited context, such as "Does self-reference lead to open-ended evolution", "How do reversible rules contribute to open-ended evolution", and "How can open-ended evolution be measured in a closed, dynamical system?"

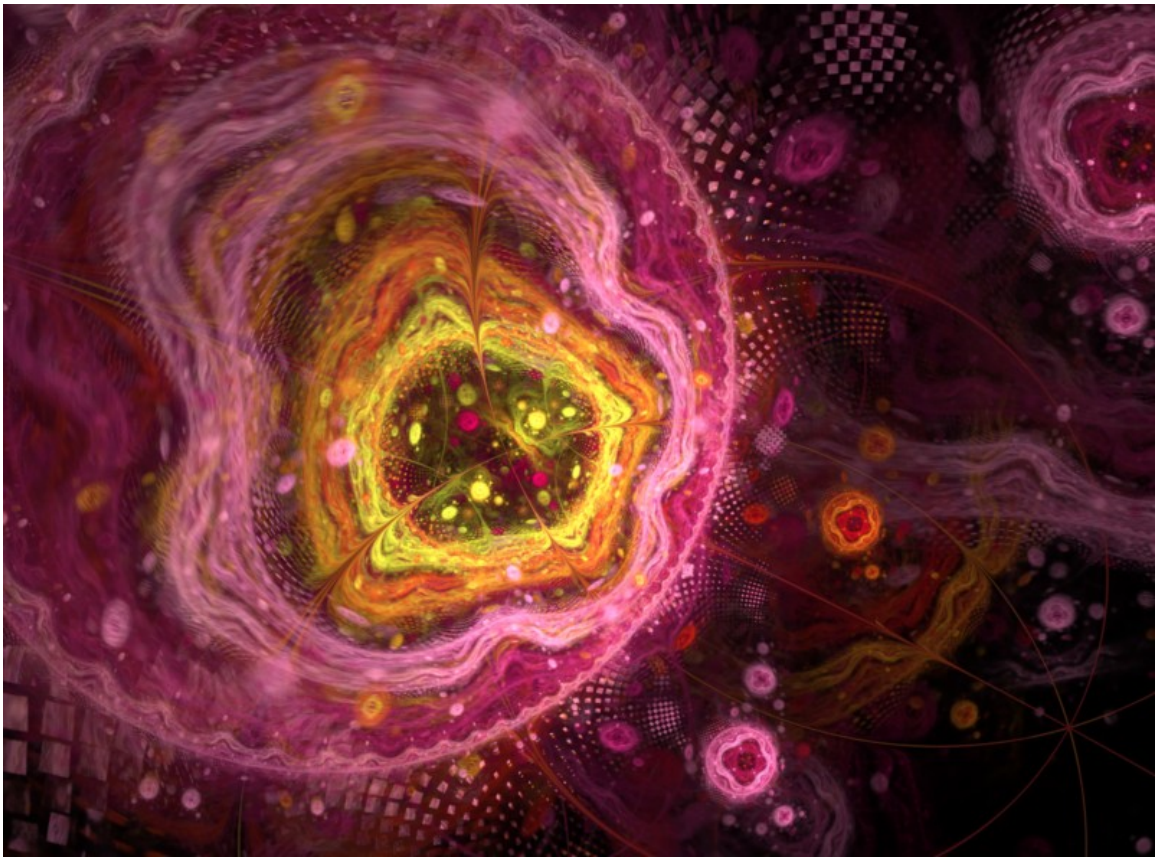


Figure 16: Artwork by Scott Draves and the Electric Sheep (Draves 2017).

2.2 A Model for Testing Open-Ended Evolution

Clearly, open-ended evolution (OEE) is the main theme of these questions. In a very general sense, this is the ability for a system to continually adapt and innovate without coming to a stopping or static point. However, OEE has been a very difficult concept to define quantitatively. Twice so far, there have been workshops at the every-other-year Artificial Life conference dedicated to OEE research and developments. At the 2016 conference, one important point was raised: the subjectivity of part of a whole system may be key to understanding biological evolution (Taylor et al. 2016).

This is reminiscent to the discussions on information in the last chapter and it seems to be a fruitful path of inquiry to explore.

However, it is difficult to think about biological OEE from our observations simply because we cannot observe the entire timeline of biology. How do we know evolution will never stop innovating? Though we have a pretty good fossil record, it is difficult to determine if fossil diversity is indication of true open-ended evolution because it relies on waiting to see what happens in the future. It's been useful to label different kinds of open-endedness to more precisely describe how our observations and studied models fit together (Vladar, Santos, and Szathmáry 2017). But no matter the kind of OEE, it must meet three key criteria as outlined in Banzhaf, Baumgaertner, and Beslon 2016: Variation, innovation, and emergence.

In order to analytically study the fossil records in terms of measurable OEE in Bedau et al. 1997, two measures were invented and were used to define long-term evolutionary dynamics. These dynamics fall into three distinct classes, where adaptive evolutionary activity is absent (class 1), bounded (class 2), or unbounded (class 3) (see Bedau et al. 1997 for a complete description). As a result, the presence of evolutionary activity has been proposed as a test for universal life (Bedau and Packard 1992).

A discussion about OEE would not be complete without including exciting virtual evolutionary worlds such as Chromaria (Soros and Stanley 2014), pictured in Figure 17. In Chromaria, four conditions for open-ended evolution were identified in the simulations. Unless all four of the conditions were met, the Chromaria's evolution would stagnate and stop. According to this study, only a special type of self-constructing, autonomous systems can implement open-ended evolution. This comes from phenotype-genotype decoupling, which originates from a new kind of way

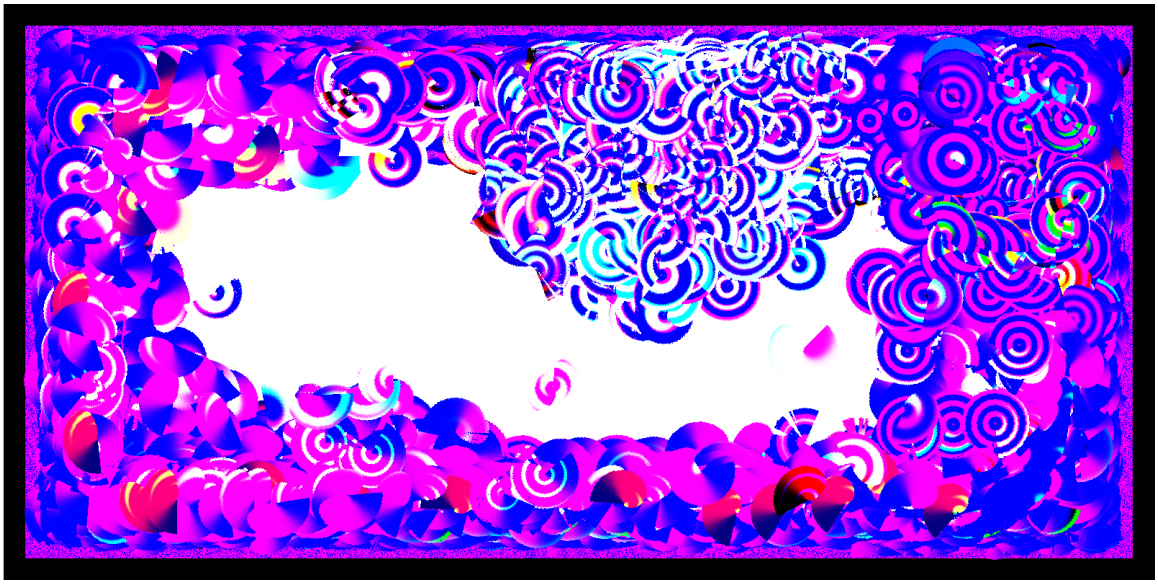


Figure 17: Screenshot of the dynamic and colorful Chromaria (image from Soros 2016).

of physically organizing materials (Ruiz-Mirazo, Umerez, and Moreno 2008), which is philosophically like the ontological principle described in Whitehead 1928.

Also, web activity (much like many other human activities) is known to exhibit open-ended evolution. Assume uploded photos are phenotypes and the annotated tags on the photos res genotypes. OEE is observed as a progressive occurence of innovative photo tags in this case (Oka, Hashimoto, and Ikegami 2015). Using phenotype-genotype analogies can be tricky and it is important to understand the mapping between phenotype and genotype objects. In collective robotic dynamics, where many robots interact with each other, OEE was observed when a "successful" genotype was allowed to spread (Bianco and Nolfi 2004). The success of a genotype depends on the phenotypic expression, such as having a genotype code that tells the robot to charge itself when its battery is low. The presence of OEE was a direct result of the genotype-phenotype map.

Although novelties are common in our everyday lives, especially with the prevalent role of the internet and technology, the mechanisms behind novelties in remain pretty much a mystery. But the states-of-the-universe analogy from Chapter 1 helps, especially when transitions between states (states) are represented as edges on a graph. In this model, the topological structure of the graph seems to change with the onset of innovation. It changes by continuously getting reshaped and expanded, resulting in distinctions between the actual and the possible (Loreto, Servedio, and Strogatz 2016). This is probably one of the greatest challenges in understanding open-ended evolution: Does biology evolve its states along a pre-defined set of possible states or does it create its own possibilities as it evolves? If the latter, then how can un-coded possible states be possible in a simple model like CA?

2.2.1 Relation to Power Laws

Power laws may or may not have some clues for understanding OEE. Zipf postulated that the distribution of word frequency in human language was a result of humans trying to communicate most efficiently with the least amount of effort possible (Zipf 1949). The distribution of the frequency of these words looks a lot like a logarithm distribution, otherwise known as a power-law distribution. Zipf's Law is illustrated in Figure 18 for several different different languages, which all follow a power law distribution. It is still unclear whether or not humans structure word frequency for the reason of communicating effectively, but there is mounting evidence that information-centered behavior (including biological behavior) can be represented with a power law distribution.

In the business world, and many other areas that use statistics, the power law

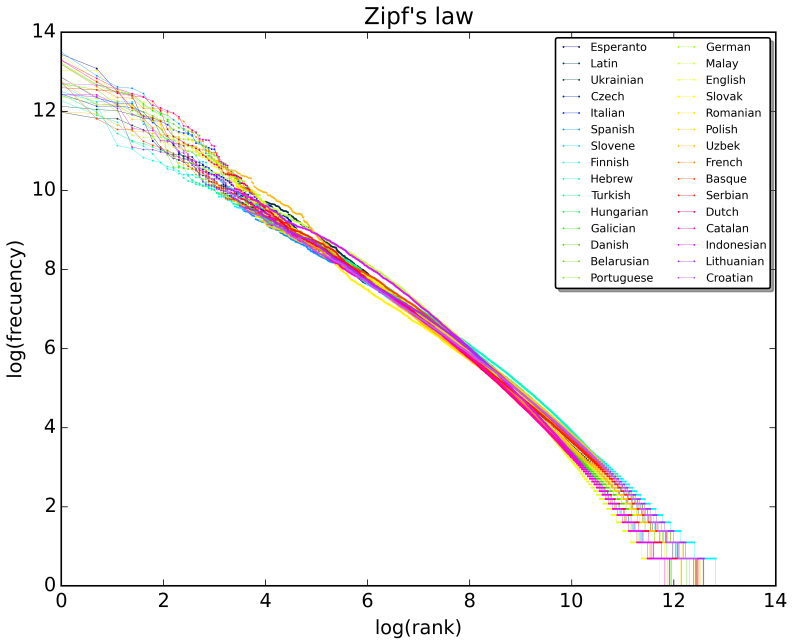


Figure 18: Specifically, the first 10 million words were taken from 30 different Wikipedias in October 2015. The frequency of the word is plotted on the y-axis with the rank of its appearence is plotted on the x-axis (image from "SergioJimenez" 2015)

is also known as the 80-20 rule (Newman 2005). It is said that 80% of sales come from 20% of customers, much like many other "ownable" assests in biological systems (Newman 2005). This makes it surprisingly easy to test if a set of words or sounds is a language or not. This is done by organizing the words/sounds according to how frequent they appear during a conversation or a large body of text. All known real languages follow the power law. This phemonenon has even been tested in other species, like zebra finches (Ma, Maat, and Gahr 2017) and has been found to be true. Could this be an easy way to test extraterrestrial signals in search of intelligent life?

The distributions of the sizes of cities, earthquakes, solar flares, moon craters, wars and people's personal fortunes all appear to follow power laws, but the origin of this behavior is still a hotly debated topic (at the time of this dissertation anyways)

Newman 2005. Many criticize the usefulness of power laws at all, stating that they could be completely meaningless. Yet others claim they are the fingerprint of open-ended systems, as in Corominas-Murtra, Seoane, and Solé 2016. In any case, the presence of a power law may indicate whether or not a system is biological or not.

Before diving down the rabbit-hole of open-ended CA systems, I should also point out that CA models which use self-reference *and* coarse-graining also exhibit power laws (Pavlic et al. 2014). In Pavlic et al. 2014, a one-dimesional CA which uses ECA rules is coarse-grained according to a particular mapping. This higher-level coarse-grained CA is 8 cells long and its state is used to determine a new ECA rule for the original CA to use for the next time step. This process is repeated at every time step. It turns out, the number of coarse-grained CA state patterns follows a power law distribution, shown in Figure 19. Either this implies coarse-graining and self-reference can lead to biological-looking systems, or nothing at all. Either way, the fact that most CA coarse-grained to a small subset of smaller CA is reminiceint of observed cellular differentiation (Pavlic et al. 2014).

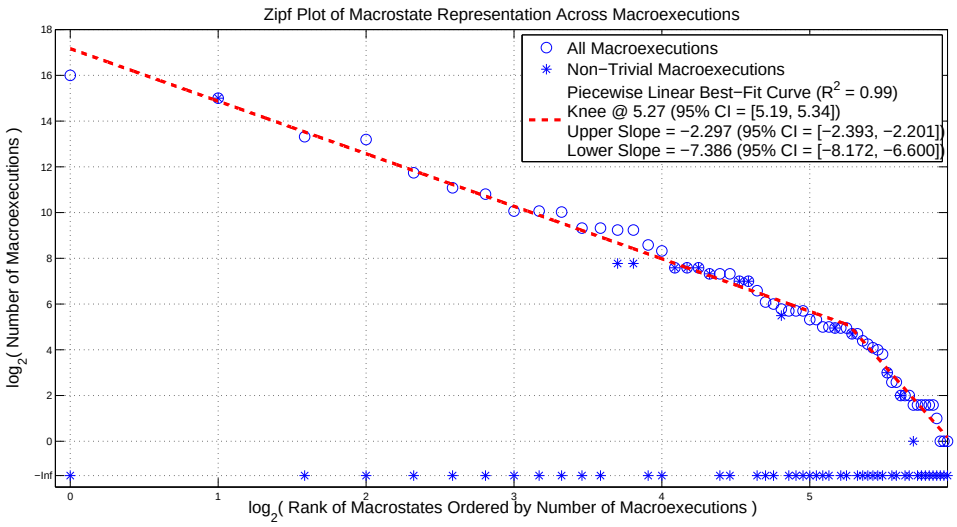


Figure 19: A Zipf plot of the number of different macrostate patterns (executions) observed in Pavlic et al. 2014. Open circles represent all macrostate executions observed while the starred points represents the data where fixed macrostate executions are removed.

Chapter 3

FORMAL DEFINITIONS OF UNBOUNDED EVOLUTION AND INNOVATION REVEAL UNIVERSAL MECHANISMS FOR OPEN-ENDED EVOLUTION IN DYNAMICAL SYSTEMS

3.1 Abstract

Open-ended evolution (OEE) is relevant to a variety of biological, artificial and technological systems, but has been challenging to reproduce *in silico*. Most theoretical efforts focus on key aspects of open-ended evolution as it appears in biology. We recast the problem as a more general one in dynamical systems theory, providing simple criteria for open-ended evolution based on two hallmark features: unbounded evolution and innovation. We define unbounded evolution as patterns that are non-repeating within the expected Poincaré recurrence time of an equivalent isolated system, and innovation as trajectories *not* observed in isolated systems. As a case study, we implement novel variants of cellular automata (CA) in which the update rules are allowed to vary with time in three alternative ways. Each is capable of generating conditions for open-ended evolution, but vary in their ability to do so. We find that *state-dependent* dynamics, widely regarded as a hallmark of life, statistically out-performs other candidate mechanisms, and is the only mechanism to produce open-ended evolution in a scalable manner, essential to the notion of ongoing evolution. This analysis suggests a new framework for unifying mechanisms for generating OEE

with features distinctive to life and its artifacts, with broad applicability to biological and artificial systems[9].

3.2 Public Summary

In this project, cellular automata (CA) were used to explore some of the mechanisms that drive open-ended evolution (OEE). In particular, two interacting CA were combined to create a system that is able to change its own laws over time. Traditional physics has fixed laws, such as the fixed Newton's Laws. They are the same with respect to time. The same idea is implemented in traditional CA; The states update according to some fixed rule (see Chapter 2 for more details on CA). Here, one of the CA in the interacting system changes its law based on the state of the system as a whole.

Figure 22 is an illustration of the interacting CA. A traditional elementary cellular automata (ECA) is on the left and our system is on the right. In the system, two separate CA interact in a few ways. One CA, called the "environment" (e), is a typical ECA and does not change its laws (or rules, in this case). The other CA is the "organism" (o) and it changes its rules based on its own current state and the current state of the environment[10].

We made the environment and organism interact three different ways:

[9]This chapter is adapted from it its original publication (A. Adams et al. 2017)

[10]Although they are called "environment" and "organism", these CA are not specifically meant to model environments and organisms (confusing, I know). Instead, they are an allegory for some external space and an internal biological-like entity, whether it be a cell or organism or species.

1. The organism changes its rule based on its own current state and the current state of the environment (Case I CA)
2. The organism changes its rule based on the environment only (Case II CA), and
3. The organism changes its rule randomly and there is no environment (Case III CA).

In these three different cases, we looked for organisms that displayed open-ended evolution. In order to be open-ended, organisms much meet two required definitions, Definitions 3.4.1 and 3.4.2. In short, these definitions state that the CA is not allowed to repeat in the same time it would have repeated using a static rule (like in an ECA) and that the CA cannot look exactly like an ECA. This ensures the organism CA is displaying novel patterns that do not repeat.

As a result, some open-ended organisms were found in all three cases but only open-ended organisms were found in Case I CA when the CA grew with size. In fact, the most amount of open-ended organisms were found when the environment is the biggest compared to the organism. As a visual example, some very large organisms are pictured in Figure 20. These are too big to determine if they meet both definitions and are indeed open-ended, but they show the idea behind innovation. Because no ECA can reproduce these patterns, they at least satisfy Definition 3.4.2.

These results are intuitive of biological systems. Since biology uses self-reference and interacting systems, it makes sense that the interacting CA model that utilizes self-reference and interactions with a second, external CA would have the most open-ended cases. While it is true that the system as a whole certainly does repeat within its expected time, this study emphasizes that quantifying open-endeness on a subjective

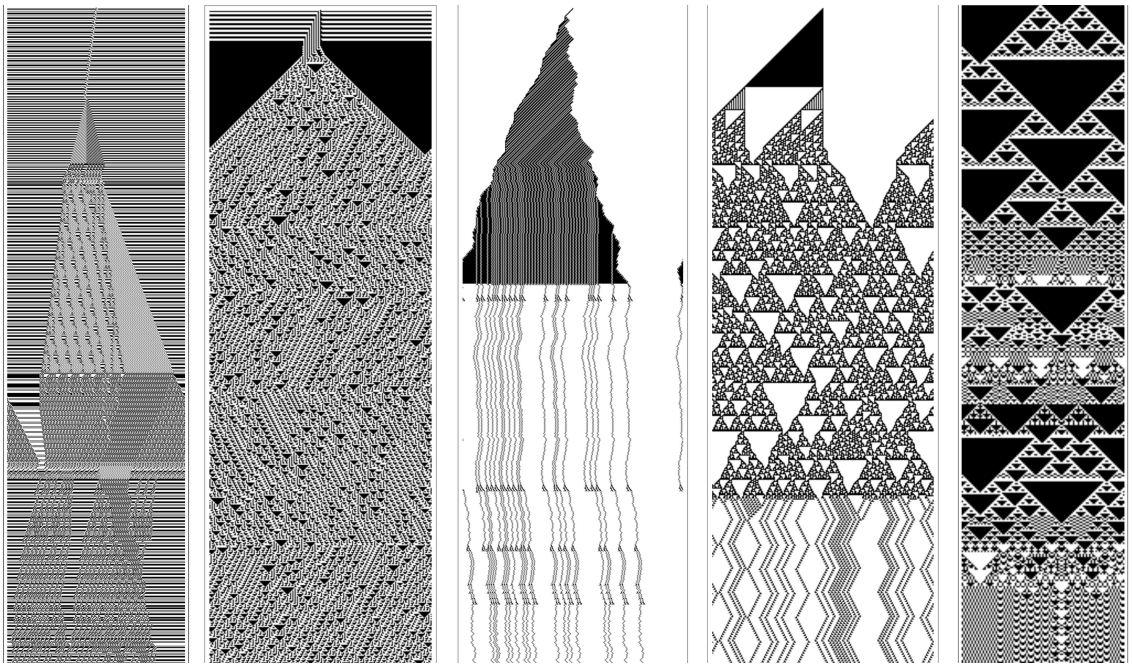


Figure 20: Some examples of large, possibly open-ended organisms.

scale (the scale of one of the CA parts) is more relevant to our understanding of biology.

3.3 Introduction

Many real-world biological and technological systems display rich dynamics, often leading to increasing complexity over time that is limited only by resource availability. A prominent example is the evolution of biological complexity: the history of life on Earth has displayed a trend of continual evolutionary adaptation and innovation, giving rise to an apparent open-ended increase in the complexity of the biosphere over its $>$ 3.5 billion year history (Bedau et al. 1997). Other complex systems, from the growth of cities (Bettencourt et al. 2007), to the evolution of language (Seyfarth, Cheney, and Bergman 2005, culture (Buchanan, Packard, and Bedau 2011

and Skusa and Bedau 2002) and the Internet (Oka, Hashimoto, and Ikegami 2015) appear to exhibit similar trends of innovation and open-ended dynamics. Producing computational models that generate sustained patterns of innovation over time is therefore an important goal in modeling complex systems as a necessary step on the path to elucidating the fundamental mechanisms driving open-ended dynamics in both natural and artificial systems. If successful, such models hold promise for new insights in diverse fields ranging from biological evolution to artificial life and artificial intelligence.

Despite the significance of realizing open-ended evolution in theoretical models, progress in this direction has been hindered by lack of a universally accepted definition for *open-ended evolution* (OEE). Although relevant to many fields, OEE is most often discussed in the context of artificial life, where the problem is so fundamental that it has been dubbed a "millennium prize problem" (Bedau et al. 2000). Many working definitions exist, which can be classified into four hallmark categories as outlined in Banzhaf, Baumgaertner, and Beslon 2016: (1) on-going innovation and generation of novelty (T. J. Taylor 1999 and Ruiz-Mirazo, Umerez, and Moreno 2008); (2) unbounded evolution (Bedau et al. 1997, and Bedau 1991, and Bedau and Packard 1992); (3) on-going production of complexity (Fernando, Kampis, and Szathmáry 2011, Ruiz-Mirazo and Moreno 2012, and Guttenberg and Goldenfeld 2008); (4) a defining feature of life (Ruiz-Mirazo, Peretó, and Moreno 2004). Each of these faces its own challenges, as each is cast in terms of equally ambiguous concepts. For example, the concepts of "innovation" or "novelty", "complexity" and "life" are all notoriously difficult to formalize in their own right. It is also not apparent whether "unbounded evolution" is physically possible since real systems are limited in their dynamics by finite resources, finite time, and finite space. A further challenge is identifying whether

the diverse concepts of OEE are driving at qualitatively different phenomena, or whether they might be unified within a common conceptual framework. For example, it has been suggested that increasing complexity might not itself be a hallmark of OEE, but instead a consequence of it (T. J. Taylor 1999 and Ruiz-Mirazo, Peretó, and Moreno 2004). Likewise, processes may appear unbounded, even within a finite space, if they can continually produce novelty within observable dynamical timescales (Taylor et al. 2016).

Given these limitations, it was unclear if OEE is a property unique to life, is inclusive of its artifacts (such as technology), or if it is an even broader phenomenon that could be a universal property of certain classes of dynamical systems. Many approaches aimed at addressing the hallmarks of OEE have been inspired by biology (Taylor et al. 2016), primarily because biological evolution is the best known example of a real-world system with the potential to be truly open-ended (Bedau et al. 1997). However, as stated, other examples of potentially open-ended complex systems do exist, such as trends associated with cultural (Buchanan, Packard, and Bedau 2011 and Skusa and Bedau 2002) and technological (Bettencourt et al. 2007 and Oka, Hashimoto, and Ikegami 2015) growth, and other creative processes. Therefore, herein we set out to develop a more general framework to seek links between the four aforementioned hallmarks of OEE within dynamical systems, while remaining agnostic about their precise implementation in biology. Our motivation is to discover *universal mechanisms* that underlie OEE as it might occur both within and outside of biological evolution.

In dynamical systems theory there exists a natural bound on the complexity that can be generated by a finite deterministic process, which is given by the *Poincaré recurrence time*. Roughly, the Poincaré time is the maximal time after which any finite system returns to its initial state and its dynamical trajectory repeats. Clearly, new

dynamical patterns cannot occur past the Poincaré time if the system is isolated from external perturbations. To cast the concept of unbounded evolution firmly within dynamical systems theory, we introduce a formal minimal criteria for *unbounded evolution* (where we stress that here we mean the broader concept of dynamical evolution, not just evolution in the biological sense) in finite dynamical systems: minimally, an unbounded system is one that does *not repeat* within the expected Poincaré time. A key feature is that this definition automatically excludes finite deterministic systems unless they are open to external perturbations in some way. That is, we contend that unbounded evolution (and in turn OEE which depends on it) is only possible for subsystem interacting with an external environment. To make better contact with real-world systems, where the Poincaré time often cannot even in principle be observed, we introduce a second criteria of *innovation*. Systems satisfying the minimal definition of unbounded evolution must also satisfy a formal notion of *innovation*, where we define innovation as dynamical trajectories *not* observed in isolated, unperturbed systems. We identify innovation by comparison to counterfactual histories (those of isolated systems). Like unbounded evolution, innovation is extrinsically defined and requires interaction between at least two subsystems. A given subsystem can exhibit OEE *if only if* it is both *unbounded* and *innovative*. As we will show, utilizing this criteria for OEE allows us to evaluate candidate mechanisms for generating OEE in simple toy model dynamical systems, ones that could carry over to more realistic complex dynamical systems.

The utility of these definitions is that they provide a simple way to quantify intuition regarding hallmarks (1) and (2) of OEE for systems of *finite* size, which is applicable to *any* comparable dynamical system. They therefore provide a means to quantitatively evaluate, and therefore directly compare, different potential mechanisms

for generating OEE. We apply these definitions to test three new variants of cellular automata (CA) for their capacity to generate OEE. A key feature of the new variants introduced is their implementation of *time-dependent update rules*, which represents a radical departure from more traditional approaches to dynamical systems where the dynamical laws remain fixed. Each variant introduced differs in its relative openness to an external environment. Of the variants tested, our results indicate that systems that implement time-dependent rules that are a function of their state are statistically better at satisfying the two criteria for OEE than dynamical systems with externally driven time-dependence for their rules (that is where the rule evolution is not dependent on the state of the subsystem of interest). We show that the state-dependent systems provide a mechanism for generating OEE that includes the capacity for on-going production of novelty by coupling to larger environments. This mechanism is also scales with system size, meaning the amount of open-endedness that is generated does not drop off as the system size increases. We then explore the complexity of state-dependent systems in more depth, calculating general complexity measures including compressibility (based on LZW in (Zenil 2010)) and Lyapunov exponents. Given that state-dependent dynamics are often cited as a hallmark feature of life due to the role of self-reference in biological processes (Walker and Davies 2013, Davies and Walker 2016, Goldenfeld and Woese 2011, and Hofstadter 1979), our results provide a new connection between hallmarks (1), (2) and (4) of OEE. Our results therefore connect several hallmarks of OEE in a new framework that allows identification of mechanisms that might operate in a diverse range of dynamical systems. The framework holds promise for providing insights into universal mechanisms for generating OEE in dynamical systems, which is applicable to both biological and artificial systems.

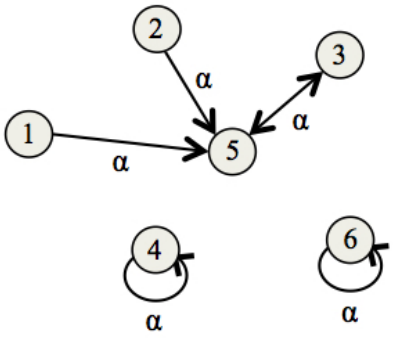
3.4 Theory

Traditionally dynamical systems, like their physical counterparts, are modeled with fixed dynamical laws – a legacy from the time of Newton. However, this framework may not be the appropriate one for modeling biological complexity, where the dynamical laws appear to be self-referential and evolve in time as a function of the states (Walker and Davies 2013, Davies and Walker 2016, Goldenfeld and Woese 2011, and Hofstadter 1979). An explicit example is the feedback between genotype and phenotype within a cell: genes are "read-out" to produce changes to the state of expressed proteins and RNAs, and these in turn can feedback to turn individual genes on and off (Noble 2012). Given this connection to biology, we are motivated in this work to focus explicitly on *time-dependent rules*, where time-dependence is introduced by driving the rule evolution through coupling to an external environment. Since open-ended evolution has been challenging to characterize in traditional models with fixed dynamical rules, implementing time-dependent rules could open new pathways to generating complexity. In this study we therefore define *open* systems as those where the rule dynamically evolves as a function of time, and we assume this is driven by interaction with an environment. As we show, time-dependent rules allow novel trajectories to be realized that have not been previously characterized in cellular automata models. To quantify this novelty, we introduce a rigorous notion of OEE that relies on formalized definitions of unbounded evolution and innovation. The definitions presented rely on utilizing isolated systems evolved according to a fixed rule as a set of counterfactual systems to compare to the novel dynamics driven by time-dependent rules.

3.4.1 Formalizing Open-Ended Evolution as Unbounded Evolution and Innovation

A hypothetical example demonstrating the concepts of INN and UE in shown in Figure 21. The possible set of states are $S = \{1, 2, 3, 4, 5, 6\}$ and rules $R = \{\alpha, \beta\}$. For each panel, the example state trajectory s is initialized with starting state $s_o = 3$. For panels **c** and **d** the rule trajectory r is also shown. Highlighted in bold is the first iteration of the attractor for states (all panels) or rules (panel **c** and **d** only). For a discrete deterministic system of six states, the Poincaré recurrence time is $t_P = 6$. Panel (**a**) shows the state transition diagram for hypothetical rule α where a trajectory initialized at $s(t_0) = 3$ visits two states. Panel (**b**) shows the state transition diagram for hypothetical rule β where a trajectory initialized at $s(t_0) = 3$ visits only one state. Since the trajectories in (**a**) and (**b**) evolve according to a fixed rule (are isolated) they do not display INN or UE and in general the recurrence time $t_r \ll t_P$. Panel (**c**) demonstrates **INN**, where the trajectory shown cannot be fully described by rule α or rule β alone. The state trajectory s and rule trajectory r both have a recurrence time of $t_r = 5$, which is less than t_P so this example does not exhibit UE. Panel (**d**) exhibits **UE** (and is also an example of INN). The trajectory shown cannot be described by rule α or rule β alone. The recurrence time for the state trajectory is $t_r = 13$, which is greater than t_P. The rule trajectory also satisfies the conditions for UE, with a recurrence time in this example that is longer than that of the state trajectory due to the fact that the state transition $2 \to 5$ could be driven by rule α or β depending on the coupling to an external system.

A hallmark feature of open-ended evolutionary systems is that they appear unbounded in their dynamical evolution (Bedau et al. 1997, Bedau 1991, and Bedau and Packard 1992). For finite systems, such as those we encounter in the real world, the

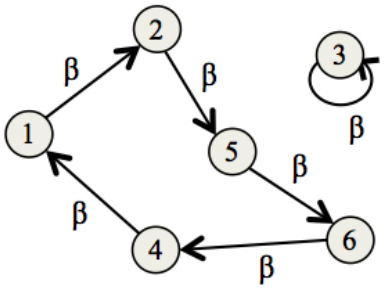


Fixed Rule α

s = {**3**, **5**, 3, 5, 3… }

r = {**α**, α, α, α, α ... }

(a)

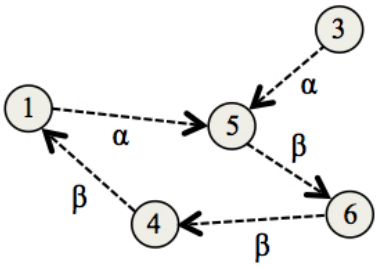


Fixed Rule β

s = {**3**, 3, 3, 3, 3… }

r = {**β**, β, β, β, β ... }

(b)

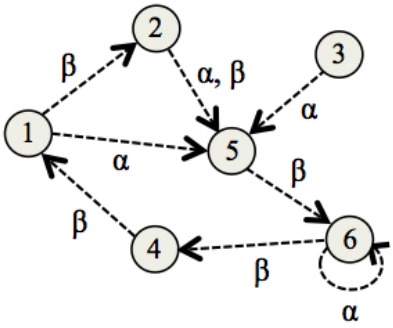


INN

s = {3, **5**, **6**, **4**, **1**, 5, 6, 4, 1… }

r = {α, **β**, **β**, **β**, **α**, β, β, β, α ... }

(c)



UE

s = {3, 5, 6, 6, **4**, **1**, **2**, **5**, **6**, **4**, **1**, **5**, **6**, 4, 1… }

r = {α, β, α, β, **β**, **β**, **α**, **β**, **β**, **β**, **β**, **β**, **β**, **α**... }

(d)

Figure 21: State diagrams of a hypothetical example demonstrating the concepts of INN and UE.

concept of "unbounded" is not well-defined. In part this is because all finite systems will eventually repeat, as captured by the well-known Poincaré recurrence theorem. As stated in the theorem, finite systems are bounded by their *Poincaré recurrence time*, which is the maximal time after which a system will start repeating its prior evolution. The Poincaré recurrence time t_P of a finite, closed deterministic dynamical system therefore provides a natural bound on when one should expect such a system to stop producing novelty. In other words, t_p is an absolute upper-bound on when such a system will terminate any appearance of open-endedness.

Potentially the Poincaré recurrence theorem can *locally* be violated by a subsystem with open boundary conditions or if the subsystem is stochastic (although in the latter case the system might still be expected to approximately repeat). We therefore consider a definition of unbounded evolution applicable to *any* instance of a dynamical system that can be decomposed into two interacting subsystems. Nominally, we refer to these two interacting subsystems as the "organism" (o) and "environment" (e). We note that our framework is sufficiently general to apply to systems outside of biology: the concept of "organism" is meant only to stress that we expect this subsystem to potentially exhibit the rich dynamics intuitively anticipated of OEE when coupled to an environment (the environment, by contrast, is not expected to produce OEE behavior). The purpose of the second "environment" subsystem is to explicitly introduce external perturbations to the organism, where e is also part of the larger system under investigation and modulates the rule of o in a time-dependent manner. We therefore minimally define *unbounded evolution* (UE) as occurring when a sub-partition of a dynamical system does not repeat within its expected Poincaré recurrence time, giving the appearance of unbounded dynamics for given finite resources:

Definition 3.4.1. Unbounded evolution (UE): A system U that can be decomposed into two interacting subsystems o and e, exhibits unbounded evolution if there exists a recurrence time such that the state-trajectory or the rule-trajectory of o is non-repeating for $t_r > t_P$ or $t'_r > t_P$ respectively, where t_r is the recurrence time of the states, t'_r the recurrence time of the rules, and t_P is the Poincaré recurrence time for an equivalent isolated (non-perturbed) system o.

Since we consider o where the states *and* rules evolve in time, unbounded evolution can apply to the state or rule trajectory recurrence time and still satisfy Definition 3.4.1. That is, a dynamical system exhibits UE *if and only if* it can be partitioned such that the sequence of one of its subsystems' states *or* dynamical rules are *non-repeating* within the expected Poincaré recurrence time t_P of an equivalent isolated system. In other words, unbounded evolution is only possible in a system that is partitioned into at least two interacting subsystems. This way, one of the subsystems acts as an external driver for the rule evolution of the other subsystem, which can then be pushed past its expected maximal recurrence time, t_P. We calculate the expected t_P as that of an equivalent isolated system. By *equivalent isolated system*, we mean the set of *all* possible trajectories evolved from any initial state drawn from the same set possible of states as for o, but generated with a fixed rule, which can be any possible fixed rule. We will describe explicit examples using the Elementary Cellular Automata (ECA) rule space in Section *Model Implementation*, where the relevant set of states are those constructed from the binary alphabet $\{0, 1\}$ and the set of rules for comparison are the ECA rules. ECA are defined as 1-dimensional CA with nearest-neighbor update rules: for an ECA of width w (number of cells across), equivalent isolated systems as defined here include all trajectories evolved with *any* fixed ECA rule from *any* initial state of width w, where t_P is then $t_P = 2^w$ and $w = w_o$, where w_o is the width of o.

Implementing the above definition of UE necessarily depends on counterfactual histories of *isolated* systems (*e.g.* of ECA in our examples). These counterfactual systems cannot, by definition, generate conditions for UE. This suggests as a corollary a natural definition for innovation in terms of comparison to the same set of counterfactual histories:

Definition 3.4.2. Innovation (INN): A system U that can be decomposed into two interacting subsystems o and e exhibits innovation if there exists a recurrence time t_r such that the state-trajectory is not contained in the set of all possible state trajectories for an equivalent isolated (non-perturbed) system.

That is, a subsystem o exhibits INN by Definition 3.4.2 if its dynamics are *not* contained within the set of all possible trajectories of equivalent isolated systems. We note these definitions do not necessitate that the complexity of individual states increase with time, thus one might observe INN without a corresponding rise in complexity with time. Fig. 21 shows a conceptual illustration of both UE and INN, as presented in Definitions 3.4.1 and 3.4.2.

A motivation for including both Definitions 3.4.1 and 3.4.2 is that they encompass intuitive notions of "on-going production of novelty" (INN) and "unbounded evolution" (UE), both of which are considered important hallmarks of OEE (Banzhaf, Baumgaertner, and Beslon 2016). UE can imply INN, but INN does not likewise imply UE. It might therefore appear that UE is sufficient to characterize OEE without needing to appeal to separately defining INN. The utility of including INN in our formalism is that it allows generalization to both infinite systems where UE is not defined, and to real-world systems where UE is not physically observable (since, for example, t_P could in principle be longer than the age of the universe). For the latter, INN can be an approximation to UE, where higher values of INN indicate a system more likely to

exhibit UE. Additionally, the combination of UE and INN can be used to exclude cases that appear unbounded but are only trivially so. For example, a partition of a system evolved according to a fixed dynamical rule could in principle locally satisfy UE, but would not satisfy INN since its dynamics could be shown to be equivalent to those generated from an appropriately constructed isolated system (*e.g.* a larger ECA in our example). An example is the time evolution of ECA Rule 30 (Wolfram 2002), which is known to be a 'complex' ECA rule that continually generates novel patterns under open-boundary conditions. In cases such as this, it should be considered that it is the complexity at the open boundary of the system that is generating continual novelty and not a mechanism *internal* to the system itself. In other words, in such examples the complexity is generated by the boundary conditions. Since our biosphere has simple, relatively homogeneous boundary conditions (geochemical and radiative energy sources) the complexity of the biosphere likely arises due to internal mechanisms and is not trivially generated by the boundary conditions alone (Smith 2008. Since we aim to understand the *intrinsic* mechanisms that might drive OEE in real, finite dynamical systems, we therefore require both definitions to be satisfied for a dynamical system to exhibit non-trivial OEE.

3.4.2 Model implementation

We evaluate different mechanisms for generating OEE against Definitions 3.4.1 and 3.4.2, utilizing the rule space of Elementary Cellular Automata (ECA) as a case study. ECA are defined as nearest-neighbor 1-dimensional CA operating on the two-bit alphabet $\{0, 1\}$. There are 256 possible ECA rules, and since the rule numbering is arbitrary, we label them according to Wolfram's heuristic designation (Wolfram 2002).

Table 1: Table of terms and model parameters.

Parameter	Definition
o	Single organism execution
e	Single environment execution
s_o	state of o
r_o	rule of o
s_e	state of e
w_o	width of o
w_e	width of e
t_P	Poincaré recurrence time
t_r	Recurrence time of s_o
t'_r	Recurrence time of r_o
I	Innovation calculated as the normalized number of rule transitions
μ	Mutation threshold of Case III variant $\mu = [0, 1)$
ξ	random noise for Case III variant, $\xi = [0, 1)$
C	Compressibility
k	Lyapunov exponent

Due to their relative simplicity, ECA represent some of the most widely-studied CA, thus providing a well-characterized foundation for this study. Traditionally, ECA evolve according to a fixed dynamical rule starting from a specified initial state. As such, no isolated finite ECA can meet both of the criteria laid out in Definitions 3.4.1 and 3.4.2 as per our construction aimed at excluding trivial cases. An isolated ECA of width w will repeat its pattern of states by the Poincaré time $t_P = 2^w$ (violating Definition 3.4.1). If we instead considered a CA of width w as a subsystem of a larger ECA it would not necessarily repeat within 2^w time steps, but it would *not* be innovative (violating Definition 3.4.2). Thus, as stated, we can exclude trivial examples such as ECA Rule 30, or other unbounded but non-innovative dynamical processes, which repeatedly apply the same update rule. A list of model parameters are summarized in Table 1.

To exclude trivial unbounded cases, Definitions 3.4.1 and 3.4.2 are constructed to require that the dynamical rules themselves evolve in time. As we will show, utilizing the set of 256 possible ECA rules as the rule space for CA with *time-dependent* rules makes both UE and INN possible. Rules can be stochastically or deterministically evolved, and we explore both mechanisms here. We note that there exists a huge number of possible variants one might consider. We therefore focus on three variants that display important mechanisms implicated in generating OEE, including openness to an environment (Ruiz-Mirazo, Umerez, and Moreno 2008) (of varying degrees in all three variants), state-dependent dynamics (regarded as a hallmark feature of life (Walker and Davies 2013, Goldenfeld and Woese 2011, and Hofstadter 1979)), and stochasticity. Here openness to an environment is parameterized by the degree to which the rule evolution of o depends on the state (or rule) of o, as compared to its dependence on the state of e. Completely open systems are regarded as depending only on external factors, such that the time-dependence of the rule evolution is *only* a function of the environment. We also consider cases that are only partially open, where the rule evolution depends on both extrinsic and intrinsic factors.

3.4.2.1 Case I CA

The first variant, Case I, implements *state-dependent* update rules, such that the evolution of o depends on its own state *and* that of its environment. This is intended to provide a model that captures the hypothesized self-referential dynamics underlying biological systems (see *e.g.* Goldenfeld and Woese (Goldenfeld and Woese 2011)) while also being open to an environment (we do not consider closed self-referential systems herein as treated in Pavlic *et al* since these do not permit the possibility of

UE (Pavlic et al. 2014)). We consider two coupled subsystems o and e, where the update rule of o is state-dependent and is a function the state and rule of o, and the state of e at the same time t (thus being self-referential but also open to perturbations from an external system). That is, the update rule of o takes on the functional form $r_o(t+1) = f(s_o(t), r_o(t), s_e(t))$, where s_o and r_o are the state and rule of the organism respectively, and s_e is the state of the environment. We regard this case as only partly open to an environment since the evolution of the rule of o depends on its own state (and rule) in addition to the state of its environment. By contrast, the subsystem e is closed to external perturbation and evolves according to a *fixed* rule (such that e is an ECA). Both o and e have periodic boundary conditions (o is only open in the sense that its rule evolution is in part externally driven). A schematic illustration of the time evolution of an ECA, and the coupling between subsystems in a Case I CA is shown in Fig. 22.

To demonstrate how the organism in our example of a Case I CA changes its update rule, we provide a simple illustrative example of the particular function $f(s_o(t), r_o(t), s_e(t))$ implemented in this work (see Fig. 23 and Supplement 1.1 in Appendix B). Specifically, we utilize an update function that takes advantage of the binary representation of ECA. An example of the structure of an ECA rule is shown for Rule 30 in Fig. 22. ECA rules are structured such that each successive bit in the binary representation of the rule is the output of one of the 2^3 possible ordered sets of triplet states. The left panel of Fig. 23 shows an example of a few times steps of the evolution of an organism o of width $w_o = 4$ (right) coupled to an environment e with width $w_e = 6$ (left), where o implemented rule 30 at $t - 1$. At each time-step t the frequency of each of the 2^3 ordered triplet states (listed in the top row of Fig. 22) in the state of o is compared to the frequency of the same ordered triplet in the state of

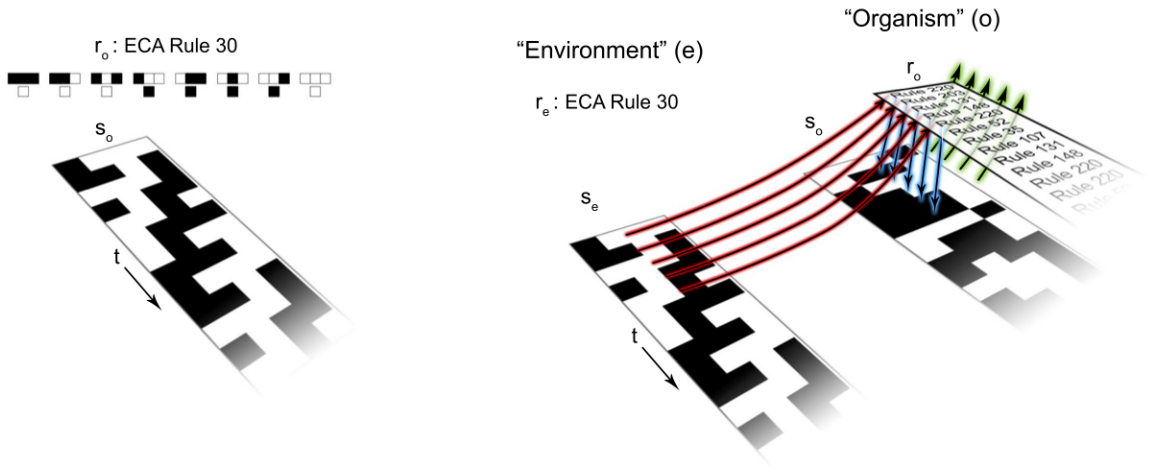
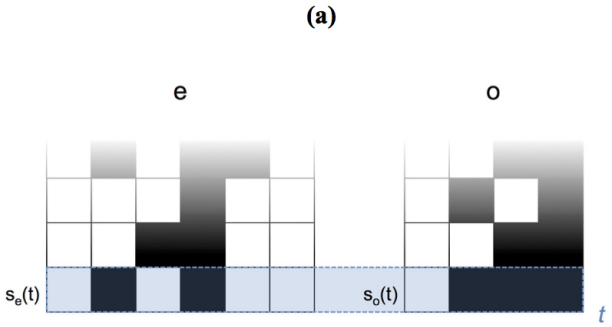


Figure 22: Illustrations of the time evolution of a standard ECA (left) and of a Case I state-dependent CA (right). ECA evolve according to a fixed update rule (here Rule 30), with the same rule implemented at each time step. In an ECA rule table, the cell representation of all possible binary ordered triplets is shown in the top row, with the cell representation of the corresponding mapping arising from Rule 30 shown below. Rule 30 therefore has the binary representation 00011110. In a Case I CA (right), the environment subsystem e evolves exactly like an ECA with a fixed rule. The organism subsystem o, by contrast, updates its rule at each time-step depending on its rule at the previous time-step, its own state (green arrows) and the state of e (red arrows). The new rule for o is then implemented to update the state of o (blue arrows). The rules are therefore time-dependent in a manner that is a function of the states of o and e and the past history of o (through the dependence on the rule at the previous time-step).

e. If the frequency in o meets or exceeds the frequency in e for a given triplet, the bit corresponding to the output of that triplet in the rule of o is flipped from $0 \leftrightarrow 1$. For the example in the left panel of Fig. 23, the triplet frequencies are listed in the table in the right panel of Fig. 23. We note that for our implementation, the frequency of a triplet in o is calculated relative to the total number of possible triplets in s_o, which is 4 in this example (and likewise for e, with 6 possible triplets in the current state). We compare the frequency only for those triplets that appear in the state of o at time t. In the table, only the triplet 101 is expressed more frequently in the organism o



(b)

triplet	frequency in $s_e(t)$	frequency in $s_o(t)$
111	-	1/4
110	-	1/4
101	1/6	1/4
100	1/6	-
011	-	1/4
010	2/6	-
001	1/6	-
000	1/6	-

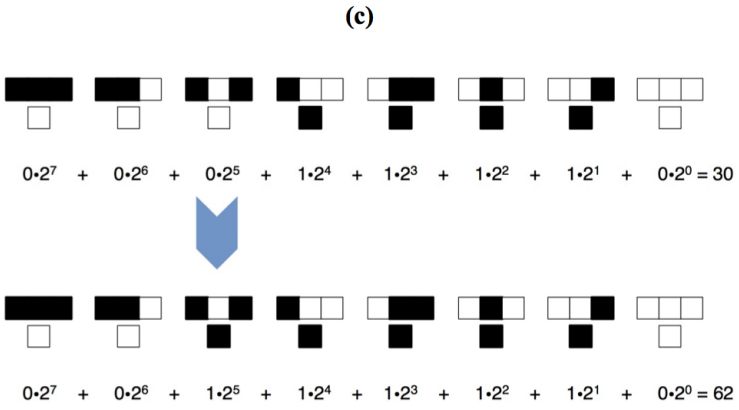


Figure 23: Example of the implementation of a Case I organism in our example. Shown is an organism o of width $w_o = 4$, coupled to an environment e with width $w_e = 6$, where the rule of o at time step t is $r_o(t) = 30$. **(a)** At each time step t, the frequency of ordered triplets are compared in the state of the organism and that of the environment, s_o and s_e respectively, and used to update $r_o(t) \to r_o(t+1)$ (see text for algorithm description). **(b)** Table of the calculated frequency of ordered triplets in the state of the environment and in the state of the organism for time step t shown in the left panel. **(c)** Update of r_o from Rule 30 to Rule 62, based on the frequency of triplets in the table **(b)**.

than in the environment e. The interaction between o and e changes r_o from Rule 30 at time-step t to Rule 62 at $t+1$, as shown in Fig. 23. The rule may change by more than one bit in its binary representation in a single time step if multiple triplets meet the criteria to change the organism's update rule.

We note that we do not expect the qualitative features of Case I CA reported here to depend on the precise form of the state-dependent update rule as presented, so long as the update of r_o depends on the state and rule of o, and the state of e (that is, o is self-referential *and* open – see Pavlic *et al.* (Pavlic et al. 2014) for an example of non-open self-referencing CA that does not display UE). We explored some variants to this rule-changing mechanism. None of our variants significantly changed the statistics of the results, indicating that the qualitative features of the dynamics do not depend on the exact (and somewhat parochial) details of the example presented herein. Instead, we regard the important part to be the general feature of self-reference coupled to openness to an environment that is driving the interesting features of the dynamics observed, where we can focus on just one example in this study for computational tractability in generating large ensemble statistics. The example implemented here was chosen since it takes explicit advantage of the structure of ECA rules (by flipping bits in the rule table) to provide a simple, open state-dependent mechanism for producing interesting dynamics.

3.4.2.2 Case II CA

We introduce a second variant of CA, Case II, that is similarly composed of two spatially segregated, fixed-width, 1-dimensional CA: an organism o and an environment e. As with Case I, the environment e is an execution of an ECA, and is evolved according to a fixed rule drawn from the set of 256 possible ECA rules. The key difference between Case I and Case II CA is that for Case II, the the update rule of the subsystem o depends *only* on the state of the external environment e and is therefore independent of the current state or rule of o – that is, o is *not* self-referential

in this example. Case II CA emulate systems where the rules for dynamical evolution are modulated exclusively by the time evolution of an external system. We consider o in this example to be more open to its environment than for Case I, since the rule evolution of o depends only on e. The functional form of Case II rule evolution may be written as $r_o(t+1) = f(s_e(t))$, where r_o is the rule of o and s_e is the state of the environment (see Supplement 1.2 in Appendix B). For the example presented here, we implement a map f that takes $s_e(t) \rightarrow r_o(t)$ that is $1:1$ from the state of e to the binary representation of the rule of o (determined according to Wolfram's binary classification scheme). Therefore for the implementation of Case II in our example the environment must be of width $w_e = 8$.

3.4.2.3 Case III CA

The final variant, Case III, is composed of a single, fixed-width, 1-dimensional CA with periodic boundary conditions, which is identified as the organism o. Like with Case II, the rule evolution of Case III is driven externally and does not depend on o. However, here the external environment e is stochastic noise and not an ECA. The subsystem o has a time-dependent rule where each bit in the rule table is flipped with a probability μ ("mutation rate") at each time step. In functional form, the subsystem o updates its rule such that $r_o(t+1) = f(r_o(t), \xi)$, where r_o is the rule of o, and ξ is a random number drawn from the interval $[0, 1)$ (see Supplement 1.3 in Appendix B). At each time step, for each bit in the rule table, a random number ξ is drawn, and if ξ is above a threshold μ, that bit is flipped $0 \leftrightarrow 1$ at that time step. This implements a diffusive-random walk through ECA rule space. Since the rule of o at time $t+1$, $r_o(t+1)$, depends on the rule at time t, $r_o(t)$, the dynamics of Case III

Table 2: Table of cellular automata variants, and the functional form of the rule evolution of o.

CA Variant	**Organism Rule Evolution**	**Environment,** e
Case I	$r_o(t+1) = f(s_o(t), r_o(t), s_e(t))$	ECA, varied w_e
Case II	$r_o(t+1) = f(s_e(t))$	ECA, $w_e = 8$
Case III	$r_o(t+1) = f(r_o(t), \xi)$	Heat bath
ECA (Isolated)	$r_o(t+1) = r_o(t)$	None

CA are history-dependent in a similar manner to Case I (both rely on flipping bits in $r_o(t)$, where Case I do so deterministically as a function of s_o and s_e, and Case III do so stochastically). In this example, o is also more open to its environment than in Case I since the organism's rule does not depend on s_o, but it is less open than Case II since the rule does depend on the previous organism rule used.

All three variants are summarized in Table 2 (see Supplement 1 in Appendix B), where the functional dependencies of the rule evolution in each example are explicitly compared. Since we restrict the rule space for Cases I–III to that of ECA rules only, the trajectories of ECA with periodic boundary conditions provides a well-defined set of isolated counterfactual trajectories with which to evaluate Definitions 3.4.1 and 3.4.2. For comparison to isolated systems, we evaluate *all* ECA of width w_o, where w_o is the width of the "organism" subsystem o. We test the capacity for each of the three cases presented to generate OEE against Definitions 3.4.1 and 3.4.2 in a statistically rigorous manner, and compare the efficacy of the different mechanisms implemented in each case.

3.4.2.4 Experimental Methods

For Cases I - III, we evolve o with periodic boundary conditions (such that interaction with the environment is only through the rule evolution). For Cases I and II, e is also a CA with periodic boundary conditions. For Case I, where w_e must also be specified, we consider systems with $w_e = 1/2w_o$, w_o, $3/2w_o$, $2w_o$ and $5/2w_o$, where w_o is the width o. For Case II, $w_e = 8$ for all simulations, since this permits a 1:1 map from the possible states of e to the rule space of ECA. Results for Case III are given for organism rule mutation rate $\mu = 0.5$, such that each outcome bit in the rule evolution has a 50% probability of flipping at every time step for ξ drawn from the interval $[0, 1)$ (a bit flips when $\mu > \xi$). Other values of μ were explored, with qualitatively similar results (see Supplement Fig. S4 in Appendix B).

The number of possible executions grows exponentially large with width w_o, limiting the computational tractability of statistically rigorous sampling. We therefore explored small CA with $w_o = 3, 4, \ldots 7$ and sampled a representative subspace of each (see Supplement 2 in Appendix B). For each system sampled, we measured the recurrence times of the rule (t'_r) and state (t_r) trajectories for o. For Case III CA, which are stochastic, all simulations eventually terminated as a random oscillation between the all '0' state and the all '1' state. We therefore used the timescale of reaching this oscillatory attractor as a proxy for the state recurrence time t_r. In cases where $t_r > t_P$ or $t'_r > t_P$, where $t_P = 2^w_o$ for isolated ECA (Definition 3.4.1), and the state trajectory was not produced by *any* ECA execution of width w_o (Definition 3.4.2), the system is considered to exhibit OEE.

We measured the complexity of the resulting interactions by calculating relative compressibility, C, and by the system's sensitivity based upon Lyapunov exponents,

k (Tisseur 2000) (see Supplement 8 in Appendix B). Large values of C indicate low Kolmogorov-Chaitin complexity, meaning the output can be produced by a simple (short) program. Large values of k indicate complex dynamics, with trajectories that rapidly diverge for small perturbations such as occurs in deterministic chaos (Tisseur 2000). These values are compared to those of ECA. Additionally, ECA rules are often categorized in terms of four Wolfram complexity classes, I - IV (Wolfram 2002). Class I and II are considered simple because all initial patterns evolve quickly into a stable or oscillating, homogeneous state. Class III and IV rules are viewed as generating more complex dynamics. We use the complexity classes of the rules utilized in time-dependent rule evolution to determine whether the complexity of time-dependent CAs is a product of the ECA rules implemented, or if it is generated through the mechanism of time-dependence.

3.5 Results

The vast majority of executions sampled from all three CA variants were innovative by Definition 3.4.2, with $> 99\%$ of Case II and Case III CAs displaying INN. For Case I CA, the percentage of INN cases increased as a function of both w_o and w_e, ranging from $\sim 30\%$ for the smallest CA explored to $> 99\%$ for larger systems (see Supplement 5 in Appendix B). This is intuitive, since the majority of organisms with changing updates rules should be expected to exhibit different state-trajectories than ECA. The fact that $> 99\%$ of organisms are innovative in our examples may seem to indicate that INN is trivial. However, we note that INN conceptually becomes more significant when considering infinite systems (where UE is not defined) or large systems where t_P is not measurable (and thus UE cannot be calculated). We show below that INN

Table 3: Percentage of sampled cases displaying OEE for each CA variant. Rows are organism width and columns correspond to the three different CA variants and ECA statistics.

w_o	**ECA**	**Case I** ($w_o = w_e$)	**Case II**	**Case III**
3	0	0.02	42.47	7.42
4	0	0.38	11.54	1.05
5	0	3.41	10.43	2.76
6	0	0.03	0.27	5.2×10^{-3}
7	0	1.06	0.7	4.7×10^{-4}

scales with recurrence time, and the amount of innovation is a good proxy for UE. INN is therefore useful to the analysis of large or infinite systems where the methods implemented here to detail candidate mechanisms are not directly applicable to test UE. INN is also necessary to exclude trivial OEE. We also note that for computational tractability we compare the time evolution of o only to ECA, but in practice one could (and perhaps should) compare o to dynamical systems evolved according to *any* fixed rule (*e.g.* regardless of neighborhood size, which for ECA is $n = 3$), in which case we might expect the number of INN cases to decrease and therefore INN would be more non-trivial even for small systems.

By contrast to cases exhibiting INN, OEE cases are much rarer, even for our highly simplified examples, due to the fact that the number of UE cases is much smaller, typically representing $< 5\%$ of all the sampled trajectories in the examples studied here. We therefore focus discussion primarily on sampled executions meeting the criteria for OEE, *i.e.* those that satisfied Definitions 3.4.1, before returning to how INN might approximate UE.

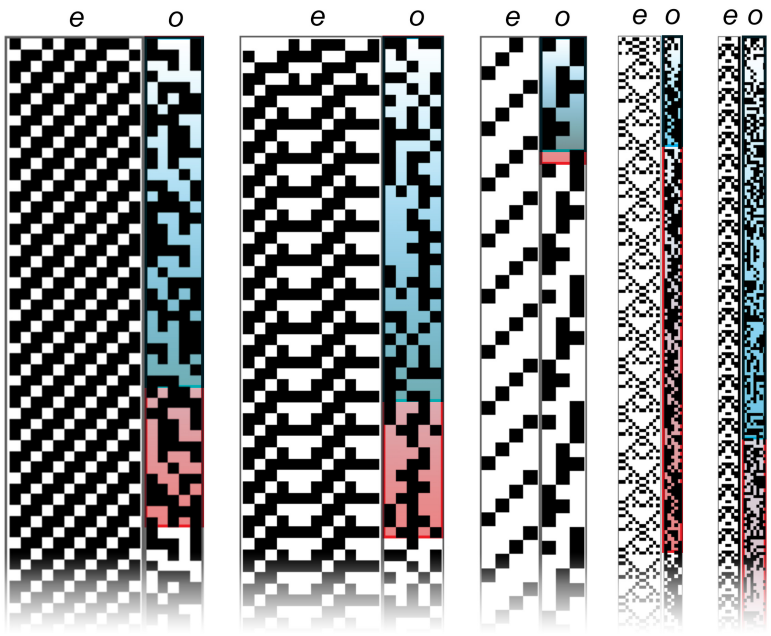


Figure 24: Examples of Case I CA exhibiting OEE. In each panel the environment e is shown on the left, and organism o on the right. For each o, the Poincaré recurrence time (t_P) for an isolated system is highlighted in blue, and the recurrence time of the states of o, t_r, is highlighted in red.

3.5.1 Open-Ended Evolution in CA variants.

The percentage of sampled cases for each CA variant that satisfy Definitions 3.4.1 for UE are shown in Table 3, where for purposes of more direct comparison Case I CA statistics are shown only for $w_o = w_e$. Case I CA statistics for other relative values of w_o and w_e are shown in Table 4. Examples of Case I CA exhibiting OEE are shown in Fig. 24, demonstrating the innovative patterns that can emerge due to time dependent rules. Box plots of the distribution of measured recurrence times for each CA variant are shown in Fig. 25 and Fig. 26. All UE cases presented here are also INN, and thus exhibit OEE. We therefore refer to UE and OEE interchangeably (without explicitly referencing OEE as cases exhibiting UE and INN separately).

To compare the capacity for OEE across the different CA variants tested, it is useful to define a notion of scalability (Taylor et al. 2016). Here we define **scalable** systems as ones where the number of observed OEE cases can increase without the need to either (1) change the rule-updating mechanism of o or (2) significantly change the statistics of sampled cases. By this definition, the two primary mechanisms for increasing the number of OEE cases in a scalable manner are by changing w_o, or depending on the nature of the coupling between o and e, changing w_e (with the constraint that the rule-updating mechanism cannot change).

As expected (by definition), isolated ECA do not exhibit any OEE cases and the majority of ECA have recurrence times $t_r \ll t_p$. However, all three CA variants with time-dependent rules do exhibit examples of OEE, but differ in the percentage of sampled cases and their scalability. Case III exhibits the simplest dynamics, where trajectories follow a diffusive random walk through rule space until the system converges on a random oscillation between the all-$'0'$ and all-$'1'$ states (where t_r is approximated by this convergence time). The frequency of state recurrence times t_r of the organism decreases exponentially (see Supplement Fig. S4 in Appendix B), such that the o with the longest recurrence times are exponentially rare. Since so few examples were found for organisms of size $w_o = 7$, we also tested $w_o = 8$ and found no examples of OEE. In general, the exponential decline observed is steeper for increasing w_o. Observing more OEE cases therefore requires exponentially increasing the number of sampled trajectories for increasing w_o. The capacity for Case III CA to demonstrate OEE is therefore not scalable with system size (violating condition 2) in our definition above). An additional limitation of Case III CA is that their the long-term dynamics are relatively simple once the system settles into the oscillatory attractor, thus the

Table 4: Percentage of sampled cases displaying OEE for Case I, with varying environment sizes.

w_o	$\frac{1}{2}w_o$	w_o	$\frac{3}{2}w_o$	$2w_o$	$\frac{5}{2}w_o$
3	0	0.02	6.52	10.81	28.14
4	0	0.38	2.28	2.94	9.65
5	0	3.41	7.04	7.5	8.64
6	0	0.03	2.15	2.64	5.82
7	0	1.06	2.95	4.39	5.34

majority of observations of Case III CA would not yield interesting dynamics (*e.g.* if the observation time were much greater than the start time $t_{obs} \gg t_o$).

For Case II, we also observed a steep decline in the number of OEE cases observed for increasing w_o (Table 3). This is reflected by a steady decrease in the mean of the recurrence times for increasing w_o, as shown in Fig. 26. We also tested a large statistical sample of organisms of size $w_o \geq 8$ for Case II CA (not shown) and found no examples of OEE cases. This is not wholly unexpected. For Case II with $w_o = 8$, the environment and organism are the same size ($w_e = w_o$). Therefore e and o share the same Poincaré time $t_P = 2^{w_o}$. The subsystem e is a traditional ECA, therefore the majority of e will exhibit recurrence times $\ll t_P$ (see *e.g.* trend in Fig. 26). Since the rule of o is determined by a 1:1 map from the state of e, the rule recurrence time of o will also be much less than the Poincaré time, such that $t'_r \ll t_P$. It is the rule evolution that drives novelty in the state evolution, we therefore also see that the state recurrence time is similarly limited such that $t_r \ll t_P$ also holds. To get around this limitation one could increase the size of the environment such that $w_e > w_o$. However, since the rule for o is a 1:1 map from the state of e, this would require changing the updating rule scheme for o. That is, the organism o would have to change how it evolves in time as a function of its environment (violating 1) in our definition of scalability above. By our definition of scalability, this is not a scalable mechanism for

generating OEE since o must change the function for its updating rule and therefore would represent a different o.

We can compare the statistics of sampled OEE cases for Case I where $w_o = w_e$ to those of Case II and Case III, as in Table 3. While Case II and Case III CA see a steep drop-off in the percentage of sampled cases exhibiting OEE with increasing organism size w_o, the Case I CA exhibit a flatter trend. We determined whether this trend holds for varying w_e by also analyzing statistics for Case I CA where $w_e = \frac{1}{2}w_o$, w_o, $\frac{3}{2}w_o$, $2w_o$ and $\frac{5}{2}w_o$. The statistics of OEE cases sampled are shown in Table 4 and box plots of the distribution of recurrence times are shown in Fig. 25. For each fixed environment size explored (w_e, columns in Table 4), we observe that the statistics do not decrease dramatically as the size of the organism increases (increasing w_o). For fixed organism size (w_o, rows in Table 4), we observe that the number of OEE cases *increases* with increasing environment size. These trends are also reflected in the means of the distributions shown in Fig. 25. Case I represents a scalable mechanism for OEE as o can be coupled to larger environments and will produce more OEE cases.

Case I and Case II can be contrasted to gain insights into scalability. The key difference between the two variants is that for Case II the update rule of o is a 1:1 map with the state of e, whereas for Case I the map is self-referencing and is *many*:1. Case I therefore uses a *coarse-grained* representation of the environment for updating the rule of o and because the dynamics are self-referential, the same pattern in the environment can lead to different rule transitions in o, depending on the previous state and rule of o. Thus, although both Case I and Case II exhibit trends of increasing OEE as w_e is increased relative to w_o, the degree to which the size of the environment can impact the time evolution of the organism is different for the two cases. For a comparable size environment in Case I and Case II CA, the pattern relevant to the

update of o may have a longer recurrence time than the actual states of e for Case I CA (due to the coarse-graining), whereas for Case II CA this pattern is strictly limited by the environment's recurrence time. Additionally, due to the coarse-graining of the environment in Case I CA, the update rule of o is not dependent on the size of e: the same exact function for updating the rule of o may be applied *independent* of the environment size. This is not true for Case II, where the function for updating the rule of o must change in order to accommodate larger environments.

3.5.2 INN as a proxy for UE

We have presented examples of small dynamical systems to perform rigorous statistical testing of INN and UE to evaluate candidate mechanisms for generating OEE. An important question is how the results might apply to larger dynamical systems that could depend on different mechanisms than those testable in simple, discrete systems. While an approximation of INN is in principle measurable for large or infinite dynamical systems, UE is not measurable or not well-defined. We therefore aimed to determine if INN can be utilized as a proxy for UE. To do so, we defined a new parameter n_r, which quantifies the number of times that an organism changed its update rule between two successive time steps in its dynamical evolution. We normalized to determine the relative **innovation** of an organism $I = \frac{n_r}{2^w}$ to generate a standardized measure for comparing across example organisms in our study. Statistically representative results for Case I and Case II organisms are shown in Fig. 27, where I is plotted against the organism's state recurrence time (Case III results are not included since the recurrence time is not well-defined). For both Case I and

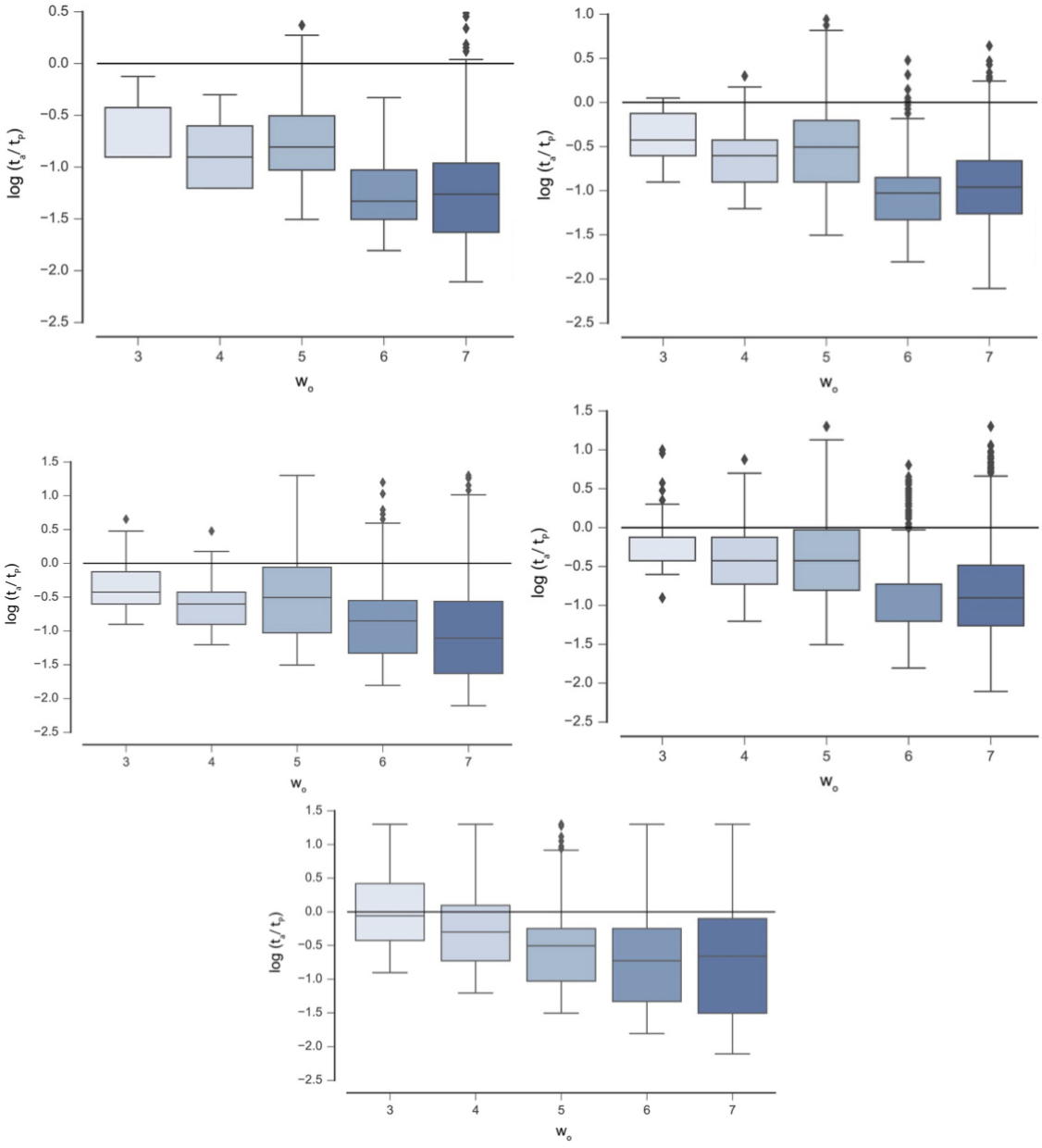


Figure 25: Distribution of recurrence times t_r for the state trajectory of o for Case I CA. From top to bottom are distributions for $w_e = 1w_o, w_o, 3w_o, 2w_o$ and $5w_o$, respectively. In all panels the black horizontal line indicates where $tr/tP = 1$ (shown on a log scale). Sampled trajectories displaying UE occur for $t_r/t_P > 1$.

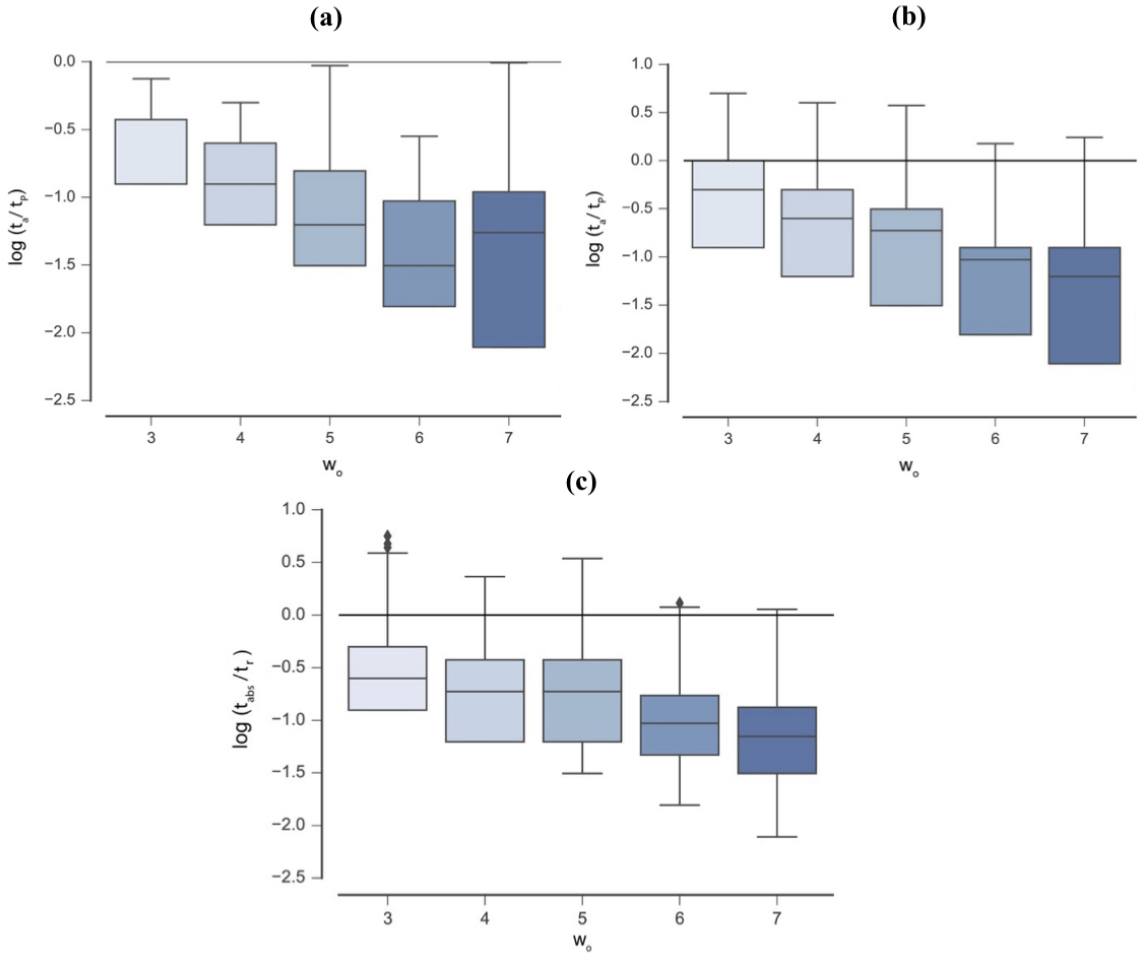


Figure 26: Distribution of recurrence times t_r for the state trajectory of o for ECA (top left), Case II (top right), and Case III CA (bottom). In all panels the black horizontal line indicates where $tr/tP = 1$ (shown on a log scale). Sampled trajectories displaying UE occur for $t_r/t_P > 1$.

Case II a clear trend is apparent where innovation is positively (and nearly linearly) correlated with recurrence time. For a given recurrence time, OEE cases (highlighted in red) are the most innovative. Comparing the two panels, it is evident that Case I CA exhibit higher innovation and therefore achieve longer recurrence times than Case II CA. From these results we can conclude that a statistical measure sampling the number of observed rule transitions could be used as a proxy for UE, which we leave as a subject for future work.

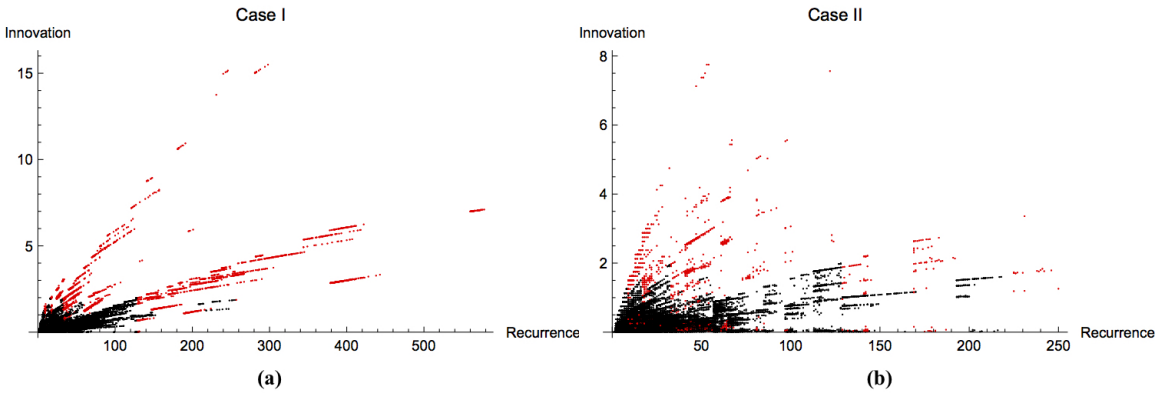


Figure 27: Relative innovation as a function of recurrence times for Case I (left) and Case II (right) CA. Highlighted in red are cases exhibiting OEE.

3.5.3 On-going Generation of Complexity in Case I

We also considered the complexity of Case I CA, relative to isolated ECA, as a further test of their scalability and potential to generate complex and novel dynamics. We characterized the complexity of Case I using two standard complexity measures, compressibility (C) and Lyapunov exponent (k). The trends demonstrate that in general C decreases with increasing organisms width w_o, but increases with increasing

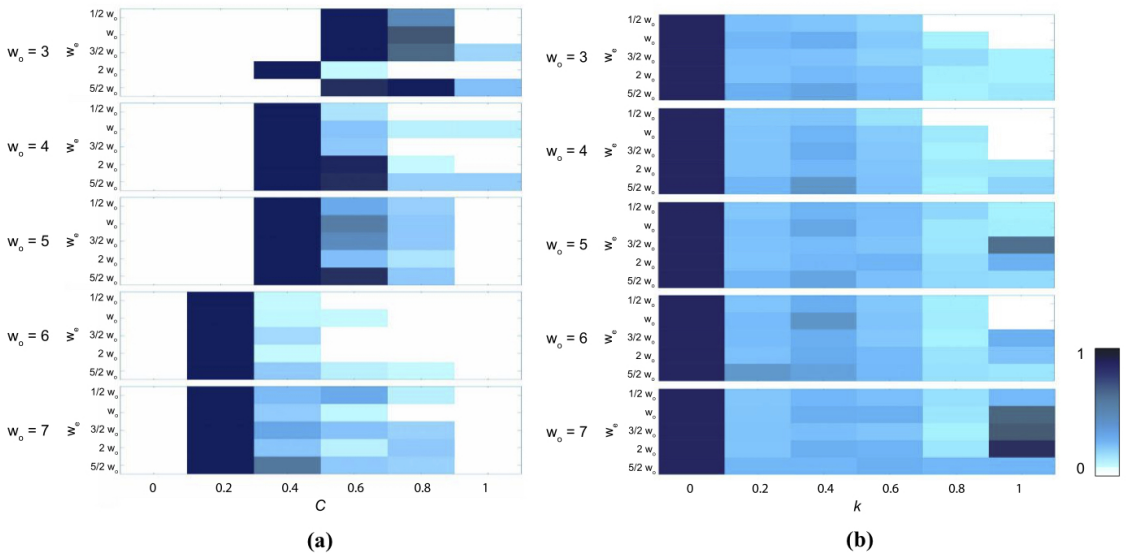


Figure 28: Heat maps of compression C (left) and Lyapunov exponent values k (right) for all state trajectories of sampled o for Case I CA. From top to bottom $w_o = 3, 4, 5, 6$ and 7, with distributions shown for $w_e = 1/2w_o$, w_o, $3/2w_o$, $2w_o$ and $5/2w_o$ (from top to bottom in each panel, respectively) for each w_o. Distributions are normalized to the total size of sampled trajectories for each w_o and w_e (see statistics in Table S3).

environment size w_e (left panel, Fig. 28), indicative of increasing complexity with organism width w_o. Similar trends are observed for the Lyapunov exponent, as shown in the right panel of Fig. 28, where it is evident that increasing w_o *or* w_e leads to an increasing number of cases with higher Lyapunov exponent k. OEE cases tend to have the highest k values (see Supplement Fig. S12 in Appendix B). As C is normalized relative to ECA (see Supplement 8 in Appendix B), we conclude that Case I CA are generally more complex than ECA evolved according to fixed dynamical rules, and this is especially true for OEE cases.

We also analyzed the ECA rules implemented in sampled Case I trajectories relative to the Wolfram Rule complexity classes. We find that Case I CA, on average, implement more Class I and II rules than Class III or IV, as shown in the frequency

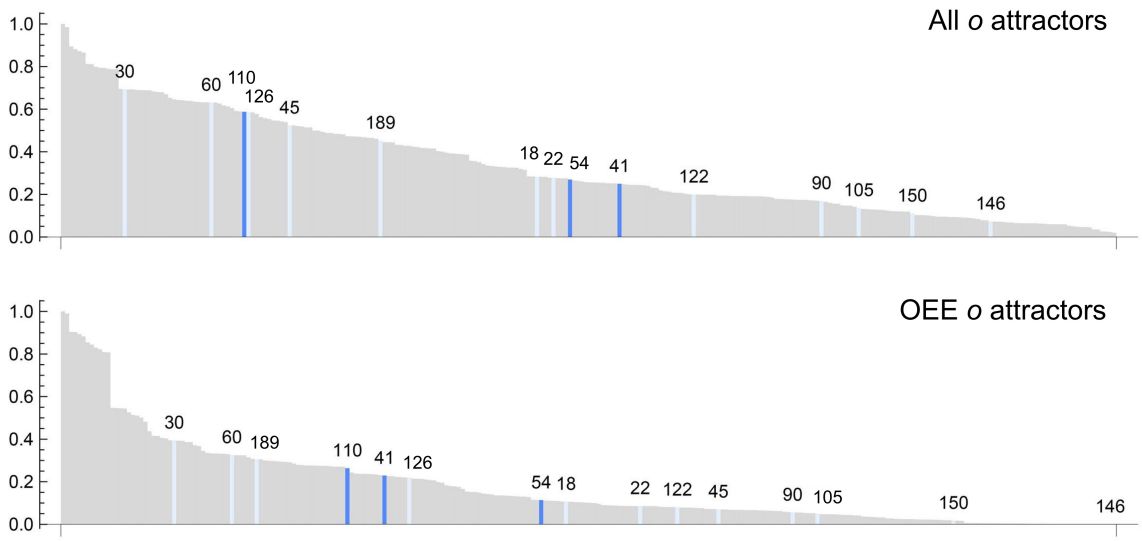


Figure 29: Rank ordered frequency distributions of rules implemented in the attractor dynamics of o for all sampled Case I CA (top) and OEE cases only (bottom). Highlighted are Wolfram Class III (light blue) and IV rules (dark blue).

distribution of Fig. 29 for Case I CA with $w_o = w_e$ (see Supplement Fig. 5 and 6 in Appendix B). Thus, we can conclude that the complexity generated by Case I CA is *intrinsic* to the state-dependent mechanism, and is not attributable to Class III and Class IV ECA rules dominating the rule evolution of o.

3.6 Discussion

We have provided formal definitions of *unbounded evolution* (UE) and *innovation* (INN) that can be evaluated in any finite dynamical system, provided it can be decomposed into two interacting subsystems o and e. Systems satisfying both UE and INN we expect to minimally represent mechanisms capable OEE. Testing the criteria for UE and INN against three different CA models with time-dependent rules reveals

what we believe to be quite general mechanisms applicable to a broad class of OEE systems.

3.6.1 Mechanisms for OEE

Our analysis indicates that there are potentially many time-dependent mechanisms that can produce OEE in a subsystem o embedded within a larger dynamical system, but that some may be more interesting than others. An externally driven time-dependence for the rules of o (Case II), while producing the highest statistics of OEE cases sampled for small o, does not provide a scalable mechanism for producing OEE with increasing system size, unless the structure of o itself is fundamentally altered (such that the rule space changes). Stochastically driven rule evolution displays rich transient dynamics, but ultimately subsystems converge on dynamics with low complexity (Case III). An alternative is to introduce stochasticity to the states, rather than the rules, which would avert this issue. This has the drawback that the mechanism for OEE is then not as clearly mappable to biological processes (or other mechanisms internal to the system), where the genotype (rules) evolve due to random mutations that then dictate the phenotype (states).

We regard Case I as the most interesting mechanism explored herein for generating conditions favoring OEE: it is scalable and the dynamics generated are novel. We note that the state-dependent mechanism represents a departure from more traditional approaches to modeling dynamical systems, *e.g.* as occurs in the physical sciences, where the dynamical rule is usually assumed to be fixed. In particular, it represents an explicit form of top-down causation, often regarded as a key mechanism in emergence (Walker and Davies 2013 and Ellis 2011) that could also play an important role in

driving major evolutionary transitions (Walker, Cisneros, and Davies 2012). The state-dependent mechanism is also consistent with an important hallmark of biology – that biological systems appear to implement self-referential dynamics such that the "laws" in biology are a function of the states (Walker and Davies 2013, Goldenfeld and Woese 2011, and Hofstadter 1979), a feature that also appears to be characteristic of the evolution of language (Levary et al. 2012 and Naoto Kataoka and Kunihiko Kaneko 2000).

3.6.2 Applicability to Other Dynamical Systems

We have independently explored openness to an environment, stochasticity and state-dependent dynamics as we expect these to be general and apply to a wide-range of dynamical systems that might similarly display OEE by satisfying Definitions 3.4.1 and 3.4.2. An important feature of these definitions is that UE and INN must be driven by *extrinsic* factors (an environment) (T. Taylor 2004), although the mechanisms driving the dynamics characteristic of OEE should be *intrinsic* to the subsystem of interest. OEE can therefore only be a property of a subsystem. We have not explored the case of feedback from o to e that might drive further open-ended dynamics, as characteristic of the biosphere, for example in niche construction (Laubichler and Renn 2015), but expect even richer dynamics to be observed in such cases. For large or infinite dynamical systems INN is an effective proxy for UE, and we expect highly innovative systems to be the most likely candidates for open-ended evolution.

3.7 Conclusions

Our results demonstrate that OEE is a general property of dynamical systems with time-dependent rules. This represents a radical departure from more traditional approaches to dynamics where the "laws" remain fixed. Our results suggest that uncovering the principles governing open-ended evolution and innovation in biological and technological systems may require removing the segregation of states and *fixed* dynamic laws characteristic of the physical sciences for the last 300 years. In particular, state-dependent dynamics have been shown to out-perform other candidate mechanisms in terms of scalability, suggestive of paths forward for understanding OEE. Our analysis connects all four hallmarks of OEE and provides a mechanism for producing OEE that is consistent with the self-referential nature of living systems. By casting the formalism of OEE within the broader context of dynamical systems theory, the proof-of-principle approach presented opens up the possibility of finding unifying principles of OEE that encompass both biological and artificial systems.

3.8 Project Acknowledgments

This project was made possible through support of a grant from the Templeton World Charity Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of Templeton World Charity Foundation. We thank S. Wolfram, T. Rowland, T. Beynon, A. Rueda-Toicen, and C. Marletto, and the Emergence group at ASU for helpful discussions related to this work.

Chapter 4

PHYSICAL UNIVERSALITY, STATE-DEPENDENT DYNAMICAL LAWS AND OPEN-ENDED NOVELTY

4.1 Abstract

A major conceptual step forward in understanding the logical architecture of living systems was advanced by von Neumann with his universal constructor, a physical device capable of self-reproduction. A necessary condition for a universal constructor to exist is that the laws of physics permit physical universality, such that *any* transformation (consistent with the laws of physics and availability of resources) can be *caused* to occur. While physical universality has been demonstrated in simple cellular automata models, so far these have not displayed a requisite feature of life – namely open-ended evolution – the explanation of which was also a prime motivator in von Neumann's formulation of a universal constructor. Current examples of physical universality rely on reversible dynamical laws, whereas it is well-known that living processes are dissipative. Here we show that physical universality and open-ended dynamics should both be possible in irreversible dynamical systems if one entertains the possibility of state-dependent laws. We demonstrate with simple toy models how the accessibility of state space can yield open-ended trajectories, defined as trajectories that do not repeat within the expected Poincaré recurrence time and are not reproducible by an isolated system. We discuss implications for physical

universality, or an approximation to it, as a foundational framework for developing a physics for life[11].

4.2 Public Summary

The model from the last section could be extended to explore other things. In particular, we wanted to see how reversible rules affect the interacting CA system. Most classic laws of physics are reversible and if biological evolution is a non-reversible process, how do the laws of physics contribute to open-ended-non-repeating-and-non-reversible evolution? We answered this question in part by using only reversible ECA rules in our previous model. The reversibility of ECA rules is a different level of reversibility of the non-repeating system. This is our analogy to the reversibility of physics on the micro-level (atoms and particles and such) and the non-reversibility of biology. The reversible ECA rules represent how the laws of physics themselves are reverible. Biology, which is instantiated in these laws of physics, is a dissipative, non-reversible process. So how do the laws of physics contribute to open-ended evolution and this irreversibility? Is it possible to bridge this conceptual gap between these two types of reversibilities?

It turns out that reversible ECA rules are surprisingly good at perpetuating open-ended organisms. When organism and environment CAs are restricted to only using reversible rules, they are much better at contributing to open-endedness and non-reversibility on the level of the organism[12]. This is likely because reversible rules

[11]This chapter is adapted from it its original publication (A. M. Adams et al. 2017)

[12]Of course the organism will repeat *eventually*, as does the entire system, but it the organism does not repeat within its Poincaré recurrence time. This is rooted in our original definition of open-ended evolution (A. Adams et al. 2017)

do not lose any information in the states, meaning the states under a reversible rule will never go to all 0's or all 1's. Reversible rules maintain the same complexity in the states over time, instead of decreasing the complexity, like non-reversible ECA rules. So if an initial state has a high amount of complexity, then it is likely to contribute to open-endedness in an organism that changes its reversible rules over time.

The biggest insight from this project is understanding how changing laws grants access to new states. We consider biology as something able to change the number of states it has access to over time, rather than evolving through a pre-defined state space. This model is capable of doing the same thing. As an organism gets a new rule, it changes which states it is able to reach in the next few time steps. Reversible rules ensure that the number of accessible states doesn't decrease over time. From this result, we were able to define local universality and provide a mechanism capable of maintaining it.

4.3 Introduction

Schrödinger's seminal 1943 lectures titled *"What is life?"* (*Schrödinger, E. "What is life?"* 1944) laid out a compelling challenge for physicists to explain the properties of living matter from what we know of physics. Based on logic and consideration of constraints imposed by physical laws (*e.g.* the second law of thermodynamics) Schrödinger was able to accurately predict that the genetic material should be "an aperiodic crystal" i.e. a stable molecule with a non-repeating pattern, which was only later discovered to be DNA. He conceded towards the end of his book that *"living matter while not eluding the 'laws of physics' as established up to date, is likely to*

involve 'other laws of physics' hitherto unknown"[13]. The challenge, as Schrödinger saw it, was to understand the function of the genetic material in purely physical terms: that is, how matter can both direct the transformations necessary for development of an organism and also how it can reliably transmit the capacity to perform those same transformations to future states of an organism (exhibiting stability) and its progeny (through self-reproduction).

Around the same time, von Neumann set out to determine the architecture of natural and artificial self-reproducing automata based again on logic combined with consideration of simple physical constraints (such as the finiteness of available resources and available time) (Von Neumann 1966). He showed that self-reproduction is logically possible for a constructor, which he defined as a machine capable of being programed to perform physical transformations, including transforming available resources to produce a copy of itself. To guarantee that self-reproduction is possible, von Neumann hypothesized the existence of a *universal constructor*: a physical system capable of being programmed to perform *any* possible physical transformation. If universal construction is possible under a given set of dynamical laws, self-reproduction is necessarily also be possible (being just one class of physical transformations). One motivation for taking this conceptual leap is to explain how biological processes could be open-ended: life on Earth is characterized by continual adaptation and innovation, giving rise to an apparent open-ended increase in the complexity of the biosphere over its several billion year history. A self-reproducing constructor is limited in its ability to reproduce by two important physical constraints[14]: the physical transformations it can

[13]Arriving at this conclusion was in fact a motivator for Schrödinger to write the book in the first place.

[14]These are in addition to more obvious universal limitations including the availability of resources and the laws of physics.

perform and the number of programs containing its specification [15]. For a universal constructor only the latter is a limiting factor, since it can perform any physical transformation (*viz* reading any possible program). Therefore, at least in principle, a universal constructor should have the maximal capacity of *any* physical device for open-ended evolution, with important implications for understanding biological evolution (Ruiz-Mirazo, Umerez, and Moreno 2008). By focusing on the logical structure of a such a system, von Neumann was able to solve part of the problem posed by Schrödinger. But he stopped short of solving the harder problem of how such a device could follow from physical principles (this harder problem was recently advanced by Marletto in (Marletto 2015) within the formalism of Constructor theory).

We do not know whether a universal constructor is itself physically possible in our universe (*The Beginning of Infinitiy: Explanations That Transform the World* 2011), or if such an entity is necessary for open-ended evolution in a dynamical system (only that it is sufficient). One necessary condition for a universal constructor to be possible is *physical universality*, defined as the property that *any* possible physical transformation can be performed on a given system, provided sufficient resources are available to do so and subject to the requirement the transformation(s) do(es) not violate any laws of physics (*e.g.* one could not build a perpetual motion machine even with a universal constructor). If physical universality can be cast as a principle of nature it could provide a promising candidate framework for arriving at the 'other laws' Schrödinger hoped might one day be uncovered. It would also likely provide new insights into the structure of physical reality, *e.g.* as is being pursued in Constructor theory, which takes as its foundation describing what transformations are possible and

[15]This includes specifications of constructors close enough to the original to permit semantic closure under mutation, see *e.g.* **Clark2017** for discussion.

why (Deutsch 2013). The key distinction between physical universality and the related, more widely discussed concept of computational universality, is that for the former computation (more aptly construction) is performed *directly* on the physical system, rather than emergent patterns (such that transformations are on states). A formal definition of physical universality was proposed by Janzig, where a dynamical system (as described by a Hamiltonian or cellular automaton) is physically universal if any transformation whatsoever can be implemented on any finite region (Janzing 2010), given sufficient resources to do so. So far three examples of physical universality have been proven in cellular automata (CA), which satisfy Janzing's definition: two classical (Schaeffer 2015 and Salo and Törmä 2017) and one quantum (Schaeffer 2015). But the proof of physically universality in each of these cases involves evolving the system to an inactive state, implying that the dynamics cannot be open-ended. Whether or not physical universality permits open-ended dynamics (such that a universal constructor, if instantiated, could exhibit open-ended evolution) remains an important open problem.

In this paper we focus on the more fundamental concept of physical universality, as a necessary prerequisite condition on the physics underlying von Neumann's universal constructor architecture (UCA). We address whether a notion of physical universality exists that is compatible with open-ended evolution, and what circumstances a system could be both physically universal and open-ended. Although our work is motivated by evolution in the biological sense, the term "evolution" intended here is broader and refers to the capacity for dynamical systems to change in time: it therefore applies both to prebiotic systems and the subsequent architectures of physical systems that have supported various stages of biological evolution. That is, we aim to address the physical architectures that permit open-ended novelty to be possible. To do so, we

investigate a new variant of cellular automata, introduced by us in (A. Adams et al. 2017 and Pavlic et al. 2014) in which the dynamics are permitted to be state-dependent. We show how state-dependent dynamics can enable both physical universality *and* open-ended evolution at the same time, by increasing the number of transformations possible in a deterministic system. In what follows, we first review the relevant properties of known physically universal CA, and explain the key differences in the model proposed herein. We then implement simple examples with state-dependent architecture to demonstrate properties of the underlying state-transition diagram which could enable physical universality and open-ended evolution. A limitation of Janzig's definition as applied to biosystems is that it necessitates dynamical laws that are reversible (such that it is possible to run them backwards in time), otherwise information in the initial state would be lost and could not be recovered by running the dynamics in reverse (that is, not every state would be accessible to the dynamics). We introduce a relaxed definition of physical universality, local physical universality, that applies to subsets of transformations, rather than all possible transformations. We show reversibility is necessary for physical universality (or its approximation) only if the dynamical laws do not evolve in time, permitting the possibility of physical universality in *irreversible* dynamical systems. Since dissipation is ubiquitous in biology, and effective descriptions of living systems are in-general state-dependent, our model provides a more appropriate starting point for understanding physical universality as it might underlie living processes, as we discuss in the conclusion.

4.4 Physical Universality and Local Physical Universality

CA are widely studied due to their simplicity and their ability to capture two key features of many real physical systems: they evolve according to a local, uniform deterministic rule, and can exhibit rich emergent behavior even from very simple rules (see *e.g.* (Israeli and Goldenfeld 2006; Hooft 2014; Wolfram 2002; Margolus 1984; Toffoli 1984; C. Langton 1990; Crutchfield 1994; Borriello and Walker 2017).[16]. For example, several CA are known to include patterns supporting computational universality (*e.g.* are Turing machines). Two well-known examples are Conway's Game of Life (Conway 1970) and ECA rule 110 (Cook 2004). In these examples, an input tape (the CA's initial state) must be formatted and typically consists of emergent spatiotemporal patterns such as gliders, particles, spaceships *etc.*. A limitation of this kind of universality is that it requires a programmer, or agent, to decide what patterns are doing the computation and how they will do it. Janzing's introduction of physical universality removes the agency in specifying computational degrees of freedom by introducing a concept of computation that can be performed arbitrarily on *all* patterns and not just the carefully constructed ones. A CA is *physically universal* if it can implement any transformation whatsoever on any finite region of the CA's state in finite time. Of note, von Neumann's original CA model of a universal constructor was not physically universal, but instead operated on patterns as in models of computational universality (first implemented by Nobili and Pesavento (Nobili and Pesavento 1994)).

[16]The physical laws governing our universe may not be completely deterministic (for example, under collapse interpretations of quantum theory) nor is reality necessarily discrete. However, by demonstrating a proof-of-principle for the more conservative and conceptually easier case of discrete deterministic systems it can be expected that at least some aspects will be sufficiently general to apply to physical laws as they might describe our real universe, under relaxed assumptions.

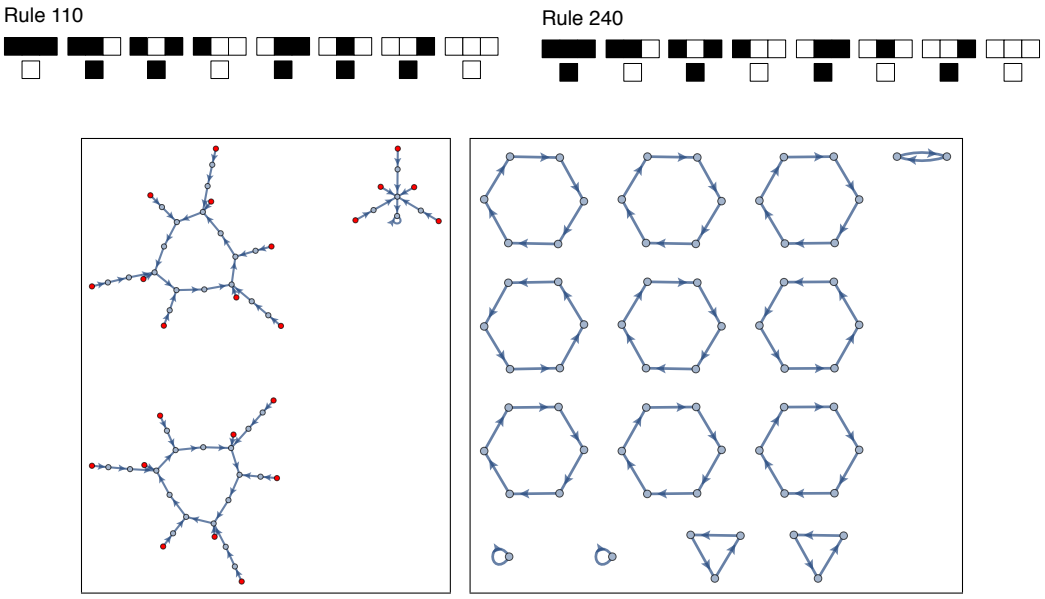


Figure 30: Examples of state-transition diagrams for $w = 6$ ECA with periodic boundary conditions. Shown are the rule table and state-transition diagram for ECA rules 110 (left) and 240 (right). Garden of Eden states are highlighted in red. Rule 240 has no garden of Eden states and is logically reversible. Each loop in the state transition diagram of rule 240 represents a set Γ of configurations that is locally physically universal (see text), whereas this property does not hold for rule 110.

Janzing's interest in developing a framework for physical universality was to provide the foundations for a physical model of control where the boundary between a controller and the physical system it controls can be shifted, and there is no difference in the physics of the controlling and controlled systems. His definition specifies two regions of the CA: a finite region (the controlled system) and the complement of the region (the surrounding cells, forming a 'controller', which is a potentially infinite resource). With this setup, physical universality is possible when any transformation of the controlled region can be implemented through the autonomous evolution of the system after the surrounding controller has been initialized in an appropriate way. (Roughly, the controlled and controller regions could be considered as data and program, respectively). More formally (adopting the slightly modified but equivalent

definition from Schaeffer 2015, and following his notation), we can define a CA where each cell has a state in the finite set Σ, with cells positioned at integer points $\mathbb{Z}^d$ in d dimensions (in this work we focus on explicit examples of Elementary Cellular Automata with $d = 1$, but consider the most general definition here). Defining $\Omega := \mathbb{Z}^d$ as the set of all cells, for a set $X \subseteq \Omega$ the compliment of X in Ω is $\bar{X} := \Omega \setminus X$ (the set of cells that *do not* belong to X in Ω). A configuration of a subset of cells $X \subseteq \Omega$ is a map assigning each cell in X a state in Σ, such that Σ^X denotes the set of all possible configurations of X. The transformation of a configuration $\gamma \in \Sigma^\Omega$ into another configuration $\gamma' \in \Sigma^\Omega$ after one time-step is denoted $\gamma \to \gamma'$. With this notation, physical universality is defined as:

Definition 4.4.1. Physical Universality: Let $X, Y \subseteq \Omega$ be finite sets. Let f be an arbitrary function that maps $\Sigma^X \to \Sigma^Y$, then a configuration $\phi \in \Sigma^{\bar{X}}$ implements the transformation f in time T if for any configuration $x \in \Sigma^X$ there exists a configuration $\psi \in \Sigma^{\bar{Y}}$ such that $x \oplus \phi T \to \psi \oplus f(x)$, where $\oplus$ is the direct sum.

A cellular automaton is *physically universal* if for any finite input region $X \subseteq \Omega$, output region $Y \subseteq \Omega$, and transformation f, there exists a configuration ϕ of $\bar{X}$ such that ϕ implements f in finite time. In other words, if a configuration ϕ of the compliment region $\bar{X}$ (the controller) can implement *any* transformation f on an input x in the region of interest, evolving it in a finite number T of steps to a configuration $f(x) \in Y$, then the CA is physically universal.

Physical universality as defined by Janzing places strict constraints on the properties of CA (or the laws of physics in the real universe). One troublesome aspect is the controller, which cannot be an emergent property of the dynamics in Janzig's formulation. Instead the controller is merely defined away as 'the rest of the universe' and can, in essence, be reduce to the laws of physics operating on an appropriate

initial state, which is not explanatory of life (see *e.g.* Sara Imari Walker 2015 for related discussion). Assuming fixed dynamical laws, one necessary (but not sufficient) condition for physical universality is reversible laws. In CA reversibility corresponds to rules that yield bijective state-transition maps from $\gamma \rightarrow \gamma'$: every state maps to exactly one other state, such that there are no "Garden of Eden" states (states that can only be an initial state). If this were not so, the CA could evolve to a configuration where no output is possible (halting the computation). Below we focus on Elementary Cellular Automata (ECA, described in Section 4.7) as the rule space for our state-dependent CA. Two examples of the state-transition diagrams for an irreversible and a reversible ECA rule for rules 110 and 240 (using Wolfram's numbering scheme in Wolfram 2002), respectively, are shown in Fig. 30 for six-bit CA ($|\Omega| = 6$) with periodic boundaries [17]. Computationally universal CA, such as rule 110, are often not reversible (an exception is Margolus's CA model of the billiard ball model of computation from Fredkin and Toffoli in Margolus 1984). Thus, even though a computationally universal CA may be able to compute any possible computable function (is computationally universal) it may not be capable of producing any possible output of cells (required for physical universality).Universal computation and universal construction are discussed by Deutsch as two separate possible principles of nature (the computability and constructibility of nature, respectively), both of which remain to be proven, see (*The Beginning of Infinitiy: Explanations That Transform the World* 2011.

Since the laws of physics are normally microscopically reversible, physical universality is at first pass compatible with known physics. However, in our universe

[17]It should be noted that for arbitrary CA specified on a Moore neighborhood (each cell is updated based on its current state and the state of its nearest neighbors) whether a CA is reversible (Myhill 1963).

there are cases where reversibility doesn't precisely hold, for example in CP violation in weak interactions. More relevant to life is the fact that, in microscopic and macroscopic physical systems, the effective laws of physics are very often irreversible, on account of the fact that they are dissipative. Irreversibility is likely essential to emergent properties in biological systems, for example, it is a requirement of formal definitions of causal emergence (Hoel 2017). Since the emergence of an arrow of time is essential to life, the question arises: *can a principle of physical universality be made compatible with a universe where irreversibility exists (if not microscopically at least as an emergent property)?*

The requirement of physical universality that every transformation be possible in a finite region is too stringent for us to make progress. However, we can consider a weaker situation in which there exist subset(s) of configurations with the property of inter-accessibility (that any transformation is possible within a given subset). We introduce the concept of *local physical universality* for this weaker case, using the following definition, directly adopted from the definition of physical universality:

Definition 4.4.2. Local Physical Universality: Let $X, Y \subseteq \Omega$ be finite sets. Let g be an arbitrary function that maps $\Gamma \to \Gamma'$, where $\Gamma \subset \Sigma^X$ and $\Gamma' \subset \Sigma^Y$. Then, a configuration $\phi \in \Sigma^{\bar{X}}$ implements the transformation g in time T if for any configuration $x \in \Gamma$ there exists a configuration $\psi \in \Sigma^{\bar{Y}}$ such that $x \oplus \phi T \to \psi \oplus g(x)$.

A cellular automaton is *locally physically universal* if for any finite input region $X \subseteq \Omega$, output region $Y \subseteq \Omega$ and transformation g, there exists a configuration of ϕ on $\bar{X}$ such that ϕ implements g in finite time.

That is, a system is locally physically universal if for any finite region $X \subset \Sigma$ all transformations are possible within a given subset of configurations Γ. Locality here is meant to indicate the connectivity of the underlying state-transition map: in a locally

physically universal cellular automata the state-transition map will contain connected components where there exists a directed path between all ordered pairs of states, defining a local neighborhood in configuration space where one can traverse a path from any state to any other state. In this sense the closed loops in the state-transition diagram of rule 240 in Fig. 30 are locally physically universal. All reversible fixed dynamical laws exhibit local physical universality. In the limit $\Gamma \to 0$, local physical universality is trivial (arbitrarily small subsets of states can be locally physically universal). In the limit $\Gamma \to \Sigma^X$ local physical universality converges to a principle of physical universality that holds over all transformations among all configurations. The utility of this definition is that it permits discussion of varying degrees of physical universality, and the possibility for the universality of a system to vary (*e.g.* as has been the case in the transition from prebiotic to biological evolution).

Schaeffer was the first to introduce an example of a physically universal CA satisfying Definition **??**. Schaeffer's CA has a local update rule that operates on a 2-by-2 block (Margolus neighborhood) in $d = 2$. The CA behaves like a diffusive gas: time evolution of a bounded region results in particles diffusing outside of that region, such that a finite region will converge to an inactive state for long times. The advantage of this is that it permits all information about what happens within a box to be intercepted as particles diffuse outside it, allowing reconstruction of what happens inside the box. Schaeffer leveraged this property to prove *any* possible transformation of particles inside a box can be programed (physical universality is possible) by suitably preparing the state of the system outside the box. Subsequently, Schaeffer demonstrated a quantum version (Schaeffer 2015) and Salo and Törma (Salo and Törmä 2017) demonstrated a one-dimensional version. Proof of physical universality for these examples follows similar logic (see Schaeffer 2015 for discussion)

and relies on tracking states backwards in time from dynamics that converge to an inactive state. Since the physically universal region in these CA terminates in a quiescent state, the physically universal region cannot exhibit another key property of life: open-ended dynamics.

4.5 State-dependent dynamical systems

We aim to show that physical universality is possible in state-dependent CA, and that for these systems open-ended evolution is also possible. State-dependent dynamical systems are ones where update rules are not fixed as a function of time, but rather change as a function of the current state of the system (for example, in biological systems when the level of gene expression in turn dictates the turning on and off of genes), such that laws and states co-evolve (Naoto Kataoka and Kunihiko Kaneko 2000, Kataoka and Kunihiko 2001, and Pavlic et al. 2014). State-dependent dynamics are frequently discussed as a hallmark feature of life (Hofstadter 1979 and Davies and Walker 2016), and provide one possible conceptual framework for understanding emergent properties in living systems in terms of physical principles (Goldenfeld and Woese 2011). In a state-dependent dynamical system, information encoded in the states determines what transformations can occur (Walker and Davies 2013). State-dependent dynamical systems therefore provide an intriguing framework for understanding physical universality in the case where some transformations are enabled by the particular information-processing architecture of physical systems.

In Adams *et al.* we showed that the co-evolution of laws and states permits greater range in the dynamical trajectories accessible to a system than what is possible with a fixed dynamical rule (A. Adams et al. 2017). The number of paths through

configuration space accessible to a system increases, since information encoded in the states enables transitions between different trajectories allowed by fixed dynamic laws (*e.g.* between the disconnected components in Fig. 30)(Pavlic et al. 2014). The minimal requirement for physical universality is that all states are a possible output of the dynamics, that is that the number of possible outputs is equivalent to the number of possible inputs of the dynamics, yielding the requirement of a bijective map for fixed rules. Because many different rules can govern transitions in a state-dependent system, the requirement is relaxed to dynamics that are surjective. This relaxed condition makes it possible, at least in principle, for a system with state-dependent rules to exhibit physical universality, even if its dynamics are *irreversible* (a surjective map is in general not reversible since it can be many-to-one or many-to-many or one-to-many).

We postulate physical universality is possible with state-dependent rules and that it can occur even in cases with irreversibility, so long as the state-transition diagram is surjective. To explore this we adapt a *state-dependent* cellular automaton introduced by us in (A. Adams et al. 2017) and show examples of systems that are locally physically universal, excluding states that contain no information (all-'0' and all-'1' states). Our original model included two interacting CA: an "organism CA" and "environment CA". To bring this abstract model in closer contact with the structure of a physical device necessary to implement such dynamics, we here recast the organism CA and environment CA as the system to be controlled and a resource (program), respectively. Our formulation does not map directly to the universal constructor architecture (UCA) proposed by von Neumann, as we are focused on the related but distinct problem of physical universality and, as noted in the introduction, do not

address the problem of biological evolution in our model (instead focusing on the prerequisite conditions for physical universality).

In von Neumann's formulation, self-reproduction of a constructor requires a physical device to store a program, which specifies the instructions for the assembly of a new copy of the constructor. Since no physical device can store the infinite amount of information necessary to specify a constructor's progeny, and all its subsequent descendants *ad infinitum* (pending available resources) a self-reproducing constructor must necessarily carry a control system – termed a supervisory unit – telling the constructor when to produce a copy of its program (or more precisely, when to assemble a duplicate of the physical system storing the program) versus when to read it. That is, the supervisory unit must decide when the instruction set is to be treated as software to be read and when it is merely hardware to be blindly copied. Drawing an analogy with this architecture, the controller (analogous to the rule of the supervisory unit) in our model is the function that specifies the interaction between the system and resource, the system plays the role of the constructor and the resource plays the role of the program or tape.

A schematic of our system is shown in Fig. 44a and a CA implementation in Fig. 44b. Both the system and resource implement ECA rules and may therefore be considered to implement the same 'physics', *e.g.* the same set of local rules. The distinction between the two emerges due to their interaction: the system's update rule is *state-dependent* and changes in time as a function of the system's state and the state of the resource CA (under direction of the meta-rule). As a concrete example of when such a setup might occur, one can consider a chemical reaction where it is desired for a particular chemical species to be transformed into another. The controller in this case could be catalyst (an example of a constructor discussed in

Deutsch 2013 and the resource is other substrates necessary for the reaction). Different transformations can be performed on the same substrate, if in contact with different catalysts and/or different resources. We additionally note the architecture we impose is consistent with other physical models of information dynamics: for example, modern incarnations of Maxwell's demon are structured such that a device is coupled to an external tape or program (see *e.g.* Mandal and Jarzynski 2012). In these models, the tape and device are not in general constrained to obey the same dynamical laws, but the thermodynamics are well defined within the context of the model (see Boyd and Crutchfield 2016 for example where demon and system obey the same dynamics). Here we do not address the thermodynamics of our devices, but within the rule structure of CA we do enforce that system and resource obey the same set of physical laws.

Janzing required that the controller and controlled system have the same physics, and that the boundary between the two systems be arbitrary. Physical universality can be demonstrated for a finite region (the controlled system) by suitably programming the initial state of its complement. The structure of our CA is different in two key aspects. The first is that the role of the controller is played by a 'metarule' that dictates the interaction between the two interacting subsystem CAs. (Note that this metarule should be constrained to obey the same laws of physics as the CA, to conform to Janzing's definition of physical universality.) We additionally define hard boundaries around the system and the resource, noting that boundaries are often discussed a fundamental to life (see *e.g.* Friston 2013). Boundaries are implemented by isolating the states of each CA with periodic boundary conditions. However, as in Janzing's case the size of the region to be programmed (the system) can be specified arbitrarily, and the only requirement is that the region be specified prior to determining if it is physically universal. Our motivation for implementing the

hard boundary is to focus on the role of state-dependent rules in enabling physical universality and open-ended evolution: we construct our system such that the *only* external interaction is through the controller's regulation of the update rule. The controller does not control the boundary of the states in our case. Because our system has closed boundary conditions, configurations containing no information – the all-'0' or all-'1' states – are cut-off from the rest of configuration space: once a system lands in one of these states it cannot escape as there is no ECA rule that can transform states outside the set of two homogeneous states. Therefore it is not possible for our system to display true physical universality; nonetheless we do show that local physical universality is possible among all other states. We expect that in cases where our devices are permitted to have boundaries that interact with other systems, physical universality will be readily possible, as in other examples of physical universality where open boundaries are necessary (such that information can flow through the boundary).

4.6 Open-ended evolution

In Adams *et al.* (A. Adams et al. 2017) we proposed formal definitions of unbounded evolution and innovation applicable to any instance of a dynamical system U that can be decomposed into two interacting subsystems $\mathcal{S}$ and $\mathcal{P}$ (such that $\mathcal{S}, \mathcal{P} \subseteq U$). These definitions are constructed to satisfy two of the four criterion for a system capable of open-ended innovation as presented by Banzhaf, Baumgaertner, and Beslon 2016, including: (1) on-going innovation and generation of novelty, (2) unbounded evolution, (3) on-going production of complexity, (4) a defining feature of life. Based criteria (1) and (2) we defined a minimal criterion for *open-ended evolution*: a region $\mathcal{X} \subseteq U$ is

open-ended if it satisfies the requirements for unbounded evolution and innovation (A. Adams et al. 2017), defined below. As we showed in A. Adams et al. 2017 the other two criteria emerge naturally as a consequence of the state-dependent framework.

Unbounded evolution (UE) can be formalized by comparing the dynamics of a system to the expected Poincaré recurrence time t_P. For any finite, deterministic dynamical system, the Poincaré recurrence time provides a natural bound on its dynamics and is defined as the timescale at which the system's dynamics should repeat if it were isolated. Unbounded evolution is therefore minimally defined as occurring only in cases where the recurrence time *locally* exceeds this bound.

Definition 4.6.1. Unbounded evolution (UE): A system U composed of at least two interacting subsystems $\mathcal{X}, \mathcal{Y} \subseteq U$ interacting according to an arbitrary function $\mathcal{F}$, exhibits unbounded evolution if there exists a recurrence time t_r such that the state trajectory $x_1 \mathcal{F} \to x_2 \mathcal{F} \to x_3 \mathcal{F} \to \ldots \mathcal{F} \to x_r$ for $x_n \in \mathcal{X}$ is non-repeating for $t_r > t_P$ and t_P is the Poincaré recurrence time $t_p = |\Sigma^{\hat{\mathcal{X}}}|$ of a finite region $\hat{\mathcal{X}}$. $\hat{\mathcal{X}}$ is an isolated equivalent of $\mathcal{X}$ without the interaction with $\mathcal{Y}$ under $\mathcal{F}$.[18]

That is, a system exhibits UE if its recurrence time t_r is greater than the Poincaré recurrence time expected for an equivalent isolated system, denoted t_P. By this definition, a system can exhibit UE *if and only if* it is interacting with an external system. Implementing this definition necessarily depends on knowledge of counterfactual histories of *isolated* systems (*e.g.* for systems implementing ECA rules, these counterfactual histories are of ECA executed with no external interaction, see (A.

[18] In (A. Adams et al. 2017 we included rule trajectories non-repeating in t_P in our definition of UE. For simplicity, since we only consider state-trajectories in the present work we do not include rule trajectories in the definition here.

Adams et al. 2017). We additionally define innovation relative to counter factual histories of isolated systems as:

Definition 4.6.2. Innovation (INN): A system U composed of at least two interacting subsystems $\mathcal{X}, \mathcal{Y} \subseteq U$ interacting according to an arbitrary function $\mathcal{F}$, exhibits innovation if there exists a recurrence time t_r such that the state trajectory $x_1 \mathcal{F}\to x_2 \mathcal{F}\to x_3 \mathcal{F}\to \ldots \mathcal{F}\to x_r$ for $x_n \in \mathcal{X}$ is *not* contained in the set of all possible state-trajectories for a finite region $\hat{\mathcal{X}}$ in isolation.

That is, a subsystem $\mathcal{X}$ exhibits innovation *if and only if* its dynamics are *not* contained within the set of all possible trajectories of equivalent isolated systems. For the case of CA, this implies that a finite region is innovative if its state trajectory cannot be reproduced by any static-rule CA with the same width w under the same rule space (here constrained to the set of ECA rules). We note that the Poincaré recurrence time for most real physical systems (including life) is much longer than the age of the universe: the utility of our definition is that it is precise and permits asking quantifiable questions about the kinds of dynamical systems that could, at least in principle, exceed this bound.

A motivation for including both Definitions 3.4.1 and 3.4.2 is that they encompass intuitive notions of "on-going production of novelty" (INN) and "unbounded evolution" (UE), both of which are considered important hallmarks of OEE, as outlined in Banzhaf, Baumgaertner, and Beslon 2016. The combination of both definitions excludes cases that continually produce complexity but are intuitively not open-ended, such as trajectories produced by ECA Rule 30, which is known to generate complex dynamics that continually create novel patterns under open-boundary conditions (Wolfram 2002). In this case, it is the open boundaries which generate the continual novelty rather than an internal mechanism of Rule 30. Since we aim to understand

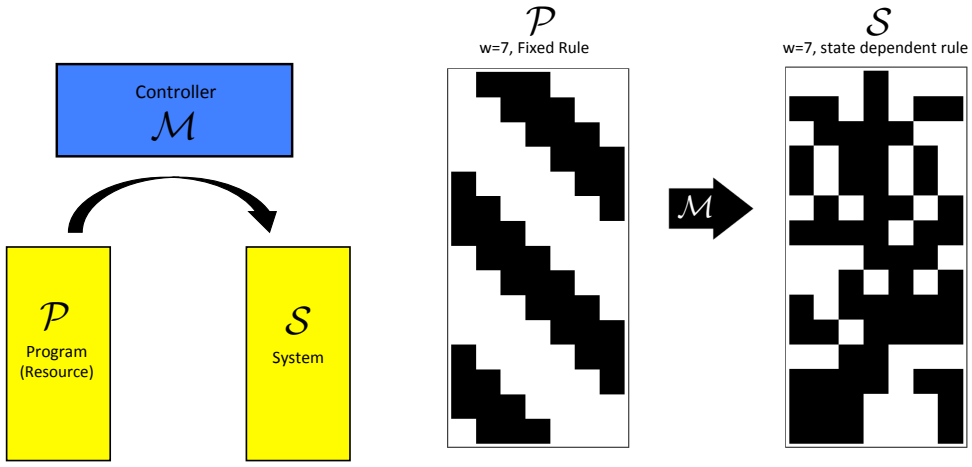


Figure 31: A state-dependent cellular automaton composed of two spatially segregated interacting parts: a system (S) is coupled to a resource program (P) through a 'metarule' M that controls the interaction. The metarule is a function that maps the states of the systems S and P to a new update rule for S.

the *intrinsic* mechanisms that might drive OEE in real, finite dynamical systems, we require that both definitions are satisfied for a dynamical system to qualify as being capable of non-trivial OEE.

Our goal here is to determine if a subsystem S can be open-ended and physically universal, or minimally at least open-ended and locally physically universal. We next introduce a more detailed description of our setup and toy model implementation and show how the capacity of state-dependent CA to satisfy the formal requirements for OEE depends on the topology of their underlying state transition diagram, focusing on whether there are regions of configuration space that are locally universal.

4.7 Methodology

Following our setup in Adams *et al* (A. Adams et al. 2017), we study a CA composed of two coupled subsystems, each using local update rules drawn from the set of Elementary Cellular Automata (ECA). ECA have $d = 1$, with nearest neighbor update rules that operate on the alphabet $\Sigma = \{0, 1\}$ (Wolfram 2002). They are popular models for studying the behavioral complexity of discrete dynamical systems. A schematic of our setup is shown in Fig. 44a and a CA implementation in Fig. 44b. The model includes the following parts:

Resource: The resource $\mathcal{P}$ is a CA of finite width w evolving according to constant function $f = r_P$, where $r_P \in R_P$ and $R_P \subseteq ECA$ is a subset of ECA rules (*e.g.* $\mathcal{P}$ is an ECA). (sets R_p used in this study are described below.)

System: The system $\mathcal{S}$ is a CA of finite width w evolving according a state-dependent map $f = r_S(t)$, where $r_S(t+1) = \mathcal{M}(\gamma_S(t), \gamma_P(t), r_S(t))$ and $\gamma_S(t)$ and $\gamma_P(t)$ are the configuration of $\mathcal{S}$ and $\mathcal{P}$ at time t, respectively, and $r_S(t)$ is the rule implemented by the system at time t. As with the resource $r_S \in R_P$ and $R_P \subseteq ECA$ is a subset of ECA rules (here the system implements an update rule drawn from the same set of rules governing the resource to ensure both obey the same "laws").

Controller: In the definition of the system, the controller $\mathcal{M}$ is an arbitrary function mapping the *rule* of the system at time t to that at $t + 1$, such that the rules of $\mathcal{S}$ evolve in time in addition to the configurations. We regard the 'metarule' $\mathcal{M}$ as the controller for the interaction between $\mathcal{P}$ and $\mathcal{S}$. For the results presented here $\mathcal{M}$ is defined as follows:

$$\mathcal{M}(t) = Round[r_S(t-1) + w\gamma_S(t) + w\gamma_P(t)] \tag{4.1}$$

where *Round* chooses the nearest upper rule r in the set of R_P, w is the width of each CA, and γ_S and γ_P are the configuration of $\mathcal{S}$ and $\mathcal{P}$ as before, respectively (these are converted from binary representation to decimal). If the expression evaluates to a value > 255 (outside the index range of ECA rules), then the evaluated value is depreciated by 255.

We studied this setup using a variety of interaction rules $\mathcal{M}$, with similar qualitative results achieved in each case. The only critical constraint on $\mathcal{M}$ is that it is be a function of $r_S(t)$, $\gamma_S(t)$ and $\gamma_P(t)$, such that the update rule of $\mathcal{S}$ depends on the previous state of the system and resource. In this way, our setup is somewhat akin to the structure of a Turing machine (or von Neumann's UCA) where the each step in a computation depends both on the internal state of the machine *and* the on the state of the tape (program) being read (Turing 1937) (that is, our system is self-referential). Because we are interested in the possibility of physical universality and open-ended dynamics in $\mathcal{S}$ we do not explicitly implement a CA instantiation of $\mathcal{M}$ here. However, we note to study physical universality (and open-endedness) as an emergent property of the dynamics one would need to explicitly implement $\mathcal{M}$ with the same sets of laws as $\mathcal{S}$ and $\mathcal{P}$. The emergence of feedback control is one of the most challenging open problems in the origins of life (Walker and Davies 2013 and Nghe et al. 2015), the exploration of which in state-dependent CA we leave as a subject of future work.

For the results presented here, both the resource and the system are implemented with periodic boundary conditions and a stationary size w (the width of the CA). As stated above, because we impose periodic boundary conditions, the examples presented cannot represent true physical universality - once an empty configuration (all '0's or all '1's) is achieved the system has no ability to generate new behavior and becomes inactive. This is also a constraint on reversible CA in the standard

approaches to physical universality, including the examples of Schaeffer and Salo and Törma (Schaeffer 2015, and Salo and Törmä 2017). It is resolved by assuming the compliment region can be made arbitrarily large, potentially requiring infinite resources. In our setup a similar solution of coupling to an arbitrarily large resource could be implemented, by permitting our system to have open boundaries (rather than periodic) in direct contact with the resource, which could be made arbitrarily large. Since our primary motivation is to isolate how state-dependent rule evolution might enable physical universality and open-ended evolution, we do not include the additional complexity of coupling the state of our system directly to the state of the resource (through a shared boundary), which will be a subject of future work. Exclusive of this constraint we do observe systems able to explore the majority of their state space, providing insights into how adopting a state-dependent framework might open new paths to understanding the ability of biological systems to explore their configuration space, in turn providing new frameworks for explaining physical universality.

4.7.1 Classifying ECA rules by reversibility

In A. Adams et al. 2017 we considered a scenario where the coupled system had access to any ECA rule for the update function of S. Here, we classify ECA rules by the fraction of all possible inputs they can produce as an output – which may roughly be considered as their *irreversibility* – and implement state-dependent CA with access to rules that share similar irreversibility. This amounts to merging the state-transition diagrams of CA with similar irreversibility to generate variants of our setup in Fig. 44b with different topologies for the underlying state-transition diagram

of $\mathcal{S}$. This permits constructing state-dependent systems with varying capacity to support the possibility of local physical universality and varying ability to support open-ended evolution. Define $\mathcal{O}$ and $\mathcal{I}$ as the set of all possible output and input states, respectively. Then, we define the "degree of irreversibility" $R_{r,w}$ of an ECA rule r for a given CA width w as the ratio of the number of unique outputs $|\mathcal{O}|$ that rule can produce (given all possible inputs $|\mathcal{I}|$) to all possible inputs $|\mathcal{I}|$, for a *single* time step t:

$$R_{r,w} = \frac{|\mathcal{O}|}{|\mathcal{I}|}. \tag{4.2}$$

For $R = 1$, a rule is fully reversible, preserving information about its past state(s). For all other values the rule is irreversible, quantified on a sliding scale by how much information the rule destroys about past states. For $R \to 0$ the system is completely irreversible and looses maximal information (*e.g.* ECA rules 0 and 255 which map all states to the all-'0' or all-'1' state, respectively). We calculated R for all ECA rules and partitioned the rules by rounding $R_{r,w}$ to the nearest 0.1, thereby identifying 11 sets R_P for each width. The relative size of each set R_P is plotted in Fig. 32 as a function of w.

By merging rules within a given reversibility class, we construct state-dependent state-transition diagrams where we can control the accessibility of states. We constructed the composition of all state-transition graphs for a given R_p, which we label as $R_p(w)$ for a given width w. Examples are shown in Fig. 33 and Fig. 34. In these diagrams, nodes represent CA states and directed edges represents a rule $r \in R_p$ that allows a transition between the connected states. The interaction function $\mathcal{M}$ and the state of the program (resource) CA together determine the realized trajectories through the state-transition diagram (and are external to this graph). For state-transition

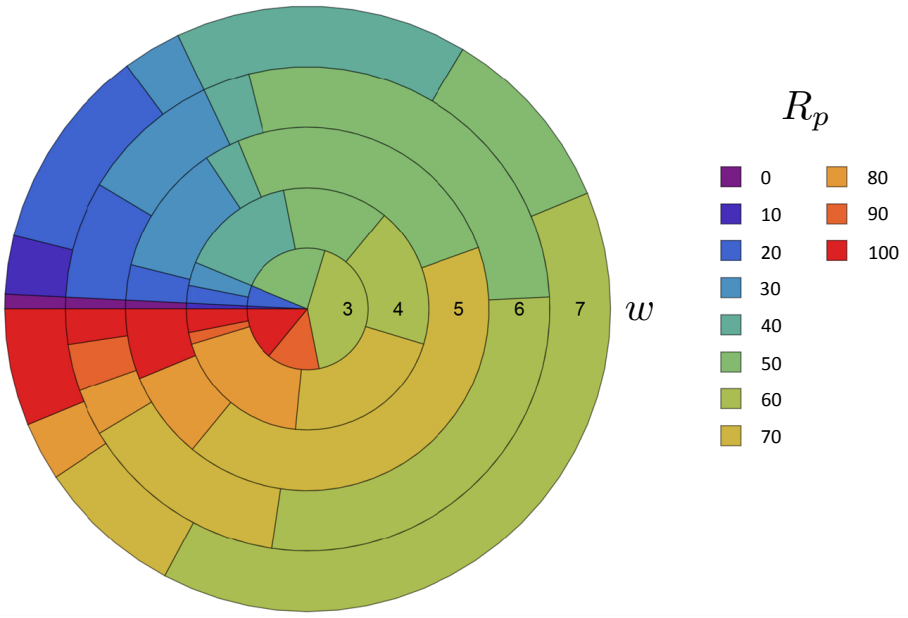


Figure 32: The relative size of all R_p sets (as percentages here) for ECA widths $w = 3, 4, 5, 6, 7$. Different w are denoted by concentric circles, where colors indicate separate R_p sets.

diagrams generated from all $R_p = 1$ rules, we know all states are accessible *a priori*. $R_p = 1$ is therefore a good candidate for studying how (local) physical universality could enable open-ended dynamics. Conversely, for $R_p = 0$ we know almost no states are accessible to the dynamics (and therefore OEE and physical universality are impossible). We expect the likelihood of open-endedness will depend in part on the topology of the state transition diagram[19]. We analyzed the structure of the state-transition diagrams using several common topological measures, including: mean in-degree, the average number of inputs for each state; .mean out-degree, the average number of outputs from each state; average shortest path, the average smallest number of transformations (steps) to move between any directed pair of states; and betweeness centrality, the number of shortest paths passing through a given state (shown in Fig.

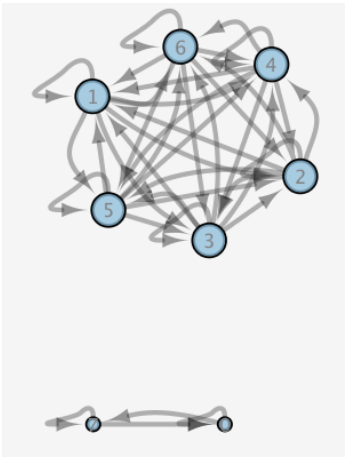
[19]There is also a resource dependence, where the recurrence time of the system in general scales linearly with that of the resource, see A. Adams et al. 2017)

33 and Fig. 34). We also determined the number of connected components and the size of the largest connected component, where a connected component is a region of the state-transition diagram where any two states are connected to each other by some (undirected) set of transformations, and which is connected to no additional states in the state-transition diagram. We equate the accessibility of state-space with physical universality if all states in a graph can be reached from any other state, the system is physically universal. If this is only true for a subset of configurations (a connected component) that set is locally physically universal. We next study the dynamics and likelihoods of generating OEE for state-transition diagrams constructed for each R_p class of a given width to determine how accessibility of the state space (topology) constrains or enables open-ended evolution.

4.8 Results

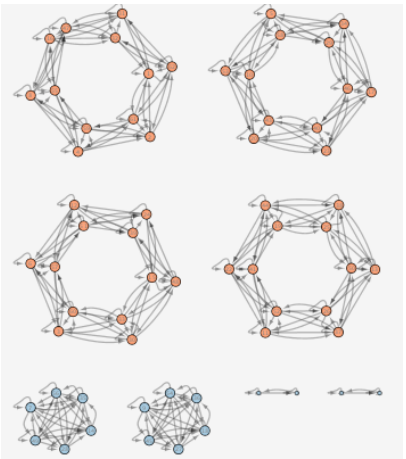
4.8.1 Likelihood of open-ended evolution

Fig. 35 shows the distributions of measured recurrence times for trajectories sampled from state-dependent CA evolved with sets of rules from each R_p classes described in Section 4.7.1, for widths $w = 3, 4, 5, 6, 7$ for $\mathcal{S}$ and $\mathcal{P}$ with $w_S = w_P$. All trajectories included in Fig. 35 are innovative, and cannot be reproduced by a static ECA rule. Trajectories above the line $t_r/t_P = 1$ in each panel are open-ended by Definitions 3.4.1 and 3.4.2. As expected, no systems with $R_p = 0$ exhibit OEE (since all information is erased and the system immediately converges to a homogeneous state). In some cases $R_p = 1$ generates the largest fraction of open-ended cases,

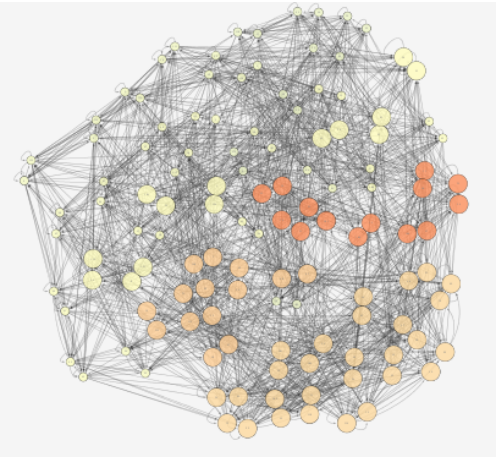


(a) $w = 3$

(b) $w = 4$

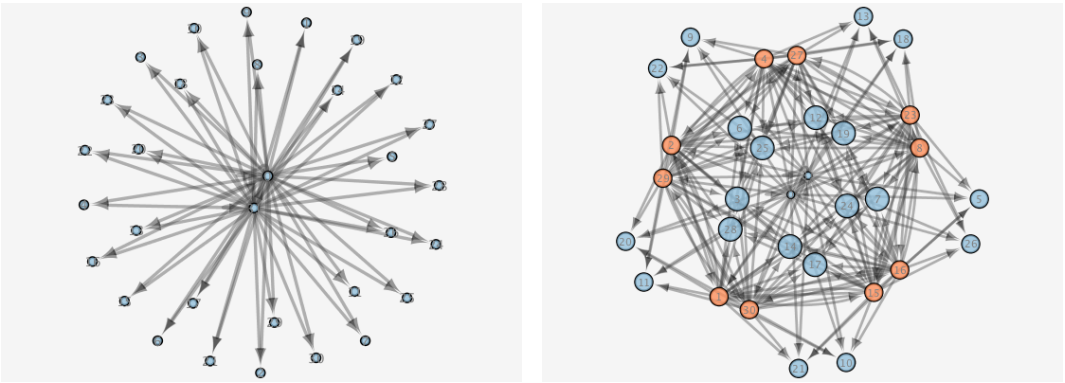


(c) $w = 5$

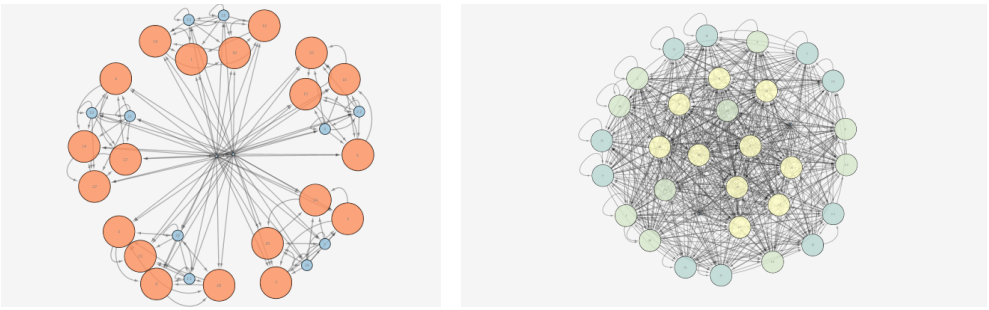


(d) $w = 7$

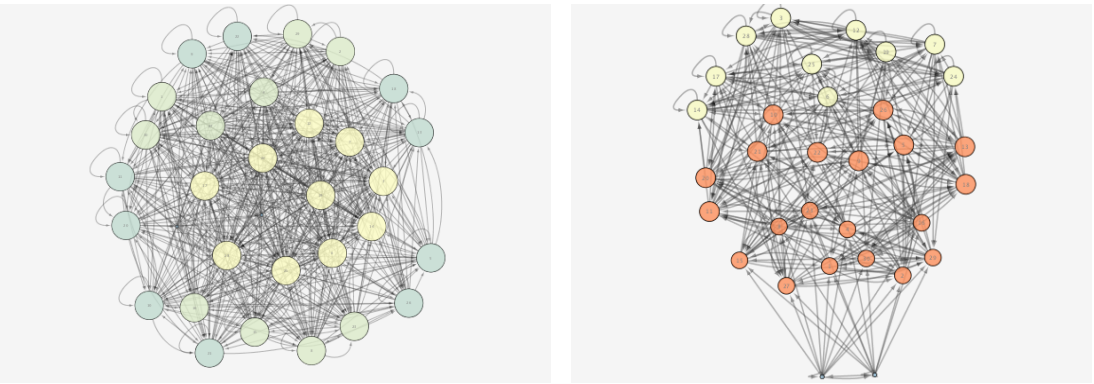
Figure 33: State transition diagrams for $R_p = 1.0$ for various w. Node size is weighted by out-degree and colors indicate betweenness centrality (high values are warm, low values cool tones). Each connected component is locally physically universal.



(a) $R_p = 0$ (b) $R_p = 20$



(c) $R_p = 40$ (d) $R_p = 50$



(e) $R_p = 70$ (f) $R_p = 80$

Figure 34: State transition diagrams for with $w = 5$ with various R_p. Node size is weighted by out-degree and colors indicate betweenness centrality (high values are warm, low values cool tones).

such as for $w = 5$ and $w = 7$. However, in other cases it is the state-transition diagrams generated from merging rules with intermediate irreversibility which exhibit the greatest capacity for producing open-ended trajectories.

4.8.2 Topology of the state-transition diagram

To better understand the patterns observed in the dynamics of state-dependent systems across different R_p classes shown in Fig. 35, we analyzed the topology of the underlying state-transition diagram for each $R_p(w)$. Analyses are shown in Table 5, including: N_rules, the number of rules in the class R_p (shown in Fig. 32); N_edges the number of edges in the state-transition graph; N_CC, the number of connected components; S_LCC the size of the largest connected component (LCC); $\langle k_in \rangle$, the mean in-degree of states (same as $\langle k_out \rangle$); $\langle l_S \rangle$, the average shortest path length between ordered pairs of states in the LCC; whether or not OEE cases were found (Y or N, respectively), and whether or not the largest connected component is locally physically universal (Y or N, respectively). Cases of interest are highlighted in red and blue, where blue denotes cases that exhibit local physical universality for the largest connected component and red denotes cases that exhibit OEE but are not locally physically universal.

If there is one component and it is locally physically universal the CA is physically universal, and any transformation is possible by suitably preparing the resource and the controller. We do not observe any such cases here for reasons stated above (states containing no information are sinks for the dynamics). However, we do see many cases of local physical universality for large connected components, which dominate the configuration space. For $R_p = 1$ all graphs are locally universal (for all fixed rules r

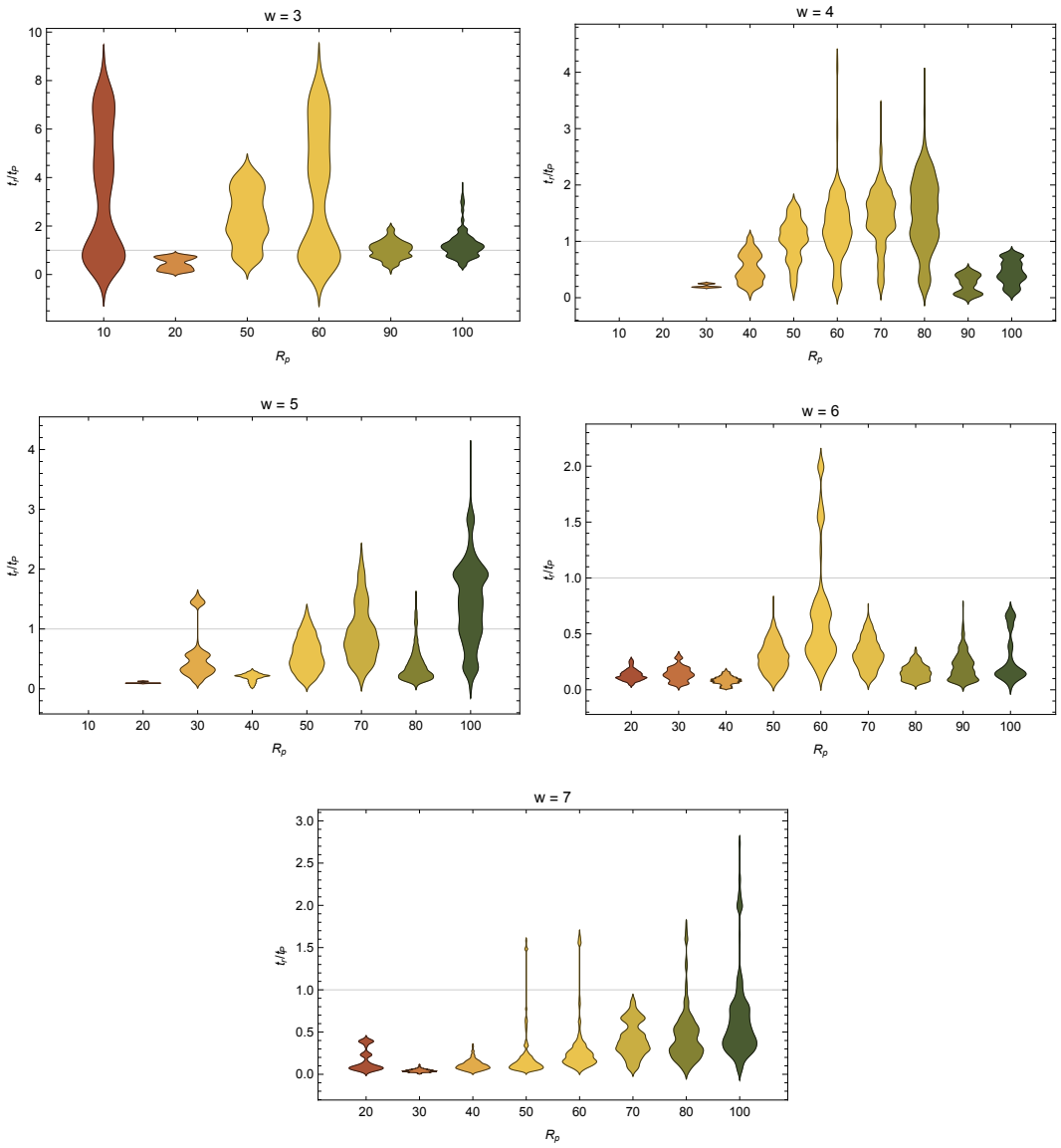


Figure 35: Distributions of recurrence times (relative to the Poincaré recurrence time, y-axis) of sampled systems S with state trajectories that are innovative, meaning they are not reproducible by any static-rule ECA. Trivial state-trajectories consisting of all 1's or 0's were removed. Distributions are shown for all R_p sets (x-axis) of a given CA widths w (panels). The horizontal line represents where $t_r = t_P$. State-trajectories above this line are classified as open-ended.

and all w if $R_{r,w} = 1$ the graph is locally physically universal, so merging graphs always will yield composite graphs that are also locally universal). In some cases, as with $w = 3, 5, 7$ merging individual rule graphs with $R_{r,w} = 1$ yields graphs with a single large connected component (excluding only the two homogeneous states), whereas in other cases the merged graph remains fragmented with many disconnected regions, *e.g.* for $w = 4, 6$[20]. OEE occurs for the former case but not the later, indicating that the larger the locally universal region in configuration space, the higher the propensity for exhibiting OEE. For other R_p, the graphs in general have a single connected component that includes all possible states, but these are not locally universal (in all cases this is in part because the all-'0' and all-'1' state are included in the graph, which are sinks for the dynamics). Nonetheless OEE is still observed in these systems. In general, more edges in the graph (more rules in the state-dependent algorithm) – corresponding to a higher mean degree and lower shortest path – are required to yield OEE for graphs that are not locally universal. That is, if the rules are not reversible, a controller must in general have a greater repertoire of rules to generate OEE than if it uses exclusively reversible rules. For two cases we see local physical universality as an emergent property of the merged graphs ($R_p = 0.9$, $w = 6$ and $R_p = 0.7$, $w = 7$), but these examples do not exhibit OEE. In both cases the graphs contain a large connected component, meaning that connectivity and the size of the connected region in configuration space together are not sufficient for OEE. None of our topological analyses identify distinctive features of these graphs. Presumably the lack of OEE in these two cases arises due to external constraints: for example resources $\mathcal{P}$ that

[20]Here, whether a merged graph composed of reversible rules is connected or not is determined by whether the CA is odd or even width, respectively, due to the structure of ECA rules and the periodic boundary conditions and is directly related to the number of rules in the set of reversible ones. For other CA, predicting whether the CA will support OEE or not will in general not be so straight forward.

terminate in fixed point attractors and thus cannot drive longer recurrence times for $\mathcal{S}$.

To further explore the relationship between topology, local universality and OEE, we also calculated the distributions of in-degree (k_{in}) and out-degree (k_{out}) for all states, where the in-degree and out-degree quantify how many transformations converge on a state or emanate from it, respectively. We also computed the shortest path l_S between all ordered pairs of states in the largest connected component, to determine the number of transformations necessary to move between any two states (for those with a path between them). Results are shown for state-transition diagrams for $w = 5$, $w = 6$ and $w = 7$ systems in Figs. 36 and 37. In general, we observe that topologies supporting open-ended evolution exhibit a long tail in the out-degree distribution, whereas those where no open-ended cases are observed exhibit a long tail in the in-degree distribution. It is difficult to be conclusive given the small networks implemented in our study, but the general trends suggest that nodes with a high in-degree act as dynamic "sinks" and those with high out-degree as "sources". In these networks high k_{out} is associated with high betweenness centrality (see *e.g.* Fig. 33 and Fig. 34, statistics not shown), which measures how many shortest paths pass through a given state. The correlation of betweeness centrality with out-degree supports the hypothesis that high-out degree states are sources for the dynamics, enabling more "short-cuts" for transformations between states. State-transition diagrams with many sources can be described as information-generating: due to their interaction with the resource more future paths are possible. By contrast sinks loose information. Physical universality requires a dynamical system where it is possible to move from any state to any other state. Our results show that a few nodes with high out-degree provide short-cuts in the dynamics and therefore may play an important role in enabling physical universality

R_p, w	N_rules	N_edges	N_{CC}	S_LCC	$\langle k_in \rangle$	$\langle l_S \rangle$	OEE	LU_{LCC}
$R_p = 0$, $w = 5$	2	64	1	100%	13	0.65	N	N
$R_p = 0.2$, $w = 5$	8	204	1	100%	41	1.35	N	N
$R_p = 0.3$, $w = 5$	30	524	1	100%	105	1.42	Y	N
$R_p = 0.4$, $w = 5$	8	164	1	100%	33	0.94	N	N
$R_p = 0.5$, $w = 5$	66	644	1	100%	129	1.3	Y	N
$R_p = 0.7$, $w = 5$	106	624	1	100%	125	1.32	Y	N
$R_p = 0.8$, $w = 5$	20	344	1	100%	69	1.65	Y	N
$R_p = 1.0$, $w = 5$	16	444	2	94%	89	1.48	Y	Y
$R_p = 0$, $w = 6$	2	128	1	100%	21	0.66	N	N
$R_p = 0.2$, $w = 6$	8	428	1	100%	71	1.41	N	N
$R_p = 0.3$, $w = 6$	36	1368	1	100%	228	1.52	N	N
$R_p = 0.4$, $w = 6$	8	420	1	100%	70	1.72	N	N
$R_p = 0.5$, $w = 6$	72	1848	1	100%	308	1.48	N	N
$R_p = 0.6$, $w = 6$	72	1548	1	100%	258	1.57	Y	N
$R_p = 0.7$, $w = 6$	36	1088	1	100%	181	1.76	N	N
$R_p = 0.8$, $w = 6$	8	428	2	97%	71	2.34	N	N
$R_p = 0.9$, $w = 6$	8	428	3	84%	71	2.46	N	Y
$R_p = 1.0$, $w = 6$	6	368	8	18%	61	1.58	N	Y
$R_p = 0$, $w = 7$	2	256	1	100%	36	0.66	N	N
$R_p = 0.1$, $w = 7$	8	900	1	100%	129	1.72	N	N
$R_p = 0.2$, $w = 7$	28	2524	1	100%	361	2.35	N	N
$R_p = 0.3$, $w = 7$	8	956	1	100%	137	1.85	N	N
$R_p = 0.4$, $w = 7$	40	3280	1	100%	469	2.14	N	N
$R_p = 0.5$, $w = 7$	26	2440	1	100%	349	2.25	Y	N
$R_p = 0.6$, $w = 7$	100	4960	1	100%	709	1.87	Y	N
$R_p = 0.7$, $w = 7$	20	1964	2	98%	281	2.26	N	Y
$R_p = 0.8$, $w = 7$	8	900	2	98%	129	3.11	Y	N
$R_p = 1.0$, $w = 7$	16	1908	2	98%	273	2.43	Y	Y

Table 5: Statistics of the topology for state-transition diagrams for S of widths $w = 5, 6, 7$. Shaded rows indicate state-transition graphs that permit OEE (red) or are that have a largest connected component that is locally universal (blue, includes open-ended cases). Included are: N_{rules}, the number of rules in the class R_p; N_{edges} the number of edges in the state-transition graph; N_{CC}, the number of connected components of the graph; S_{LCC} the size of the largest connected component (LCC); the mean in-degree of states, the average shortest path length between directed pairs of states in the LCC; whether or not OEE cases were found (Y or N, respectively), and whether or not the largest connected component is locally physically universal (Y or N, respectively).

and open-ended evolution. Future work will determine if such networks display small world properties, as is characteristic of many real-world systems (Watts and Strogatz 1998.

4.9 Conclusions

It is unknown whether physical universality is a property of our universe. Here we demonstrated a weaker form of physical universality, which we call *local physical universality*, to provide a framework for understanding cases where true physical universality is not possible or only approximately holds. This may be the case, for example, if the state space cannot be predefined, as proposed by Kauffman S. Kauffman 2000. In a system that is locally physically universal, there exists a subset of configurations among which all transformations in the set are possible. Cellular life is often described as an example of a universal constructor (see *e.g.* Danchin 2009; Hickinbotham and Stepney 2016). This kind of universality is most appropriately cast as a local one. The "universal constructor" (often approximated as the ribosome (see Walker and Davies 2013 for discussion) operates on a finite set of materials – including amino and nucleic acids among others. An example of a set of configurations which is approximately locally physically universal because of the information-processing architecture of life is the set of proteins composed of the 20 (or so) genetically encoded amino acids: in principle, given sufficient resources a biological system should be able to construct any such protein (or to convert resources from one to another). This example highlights how, when it comes to real physical systems, such as life,

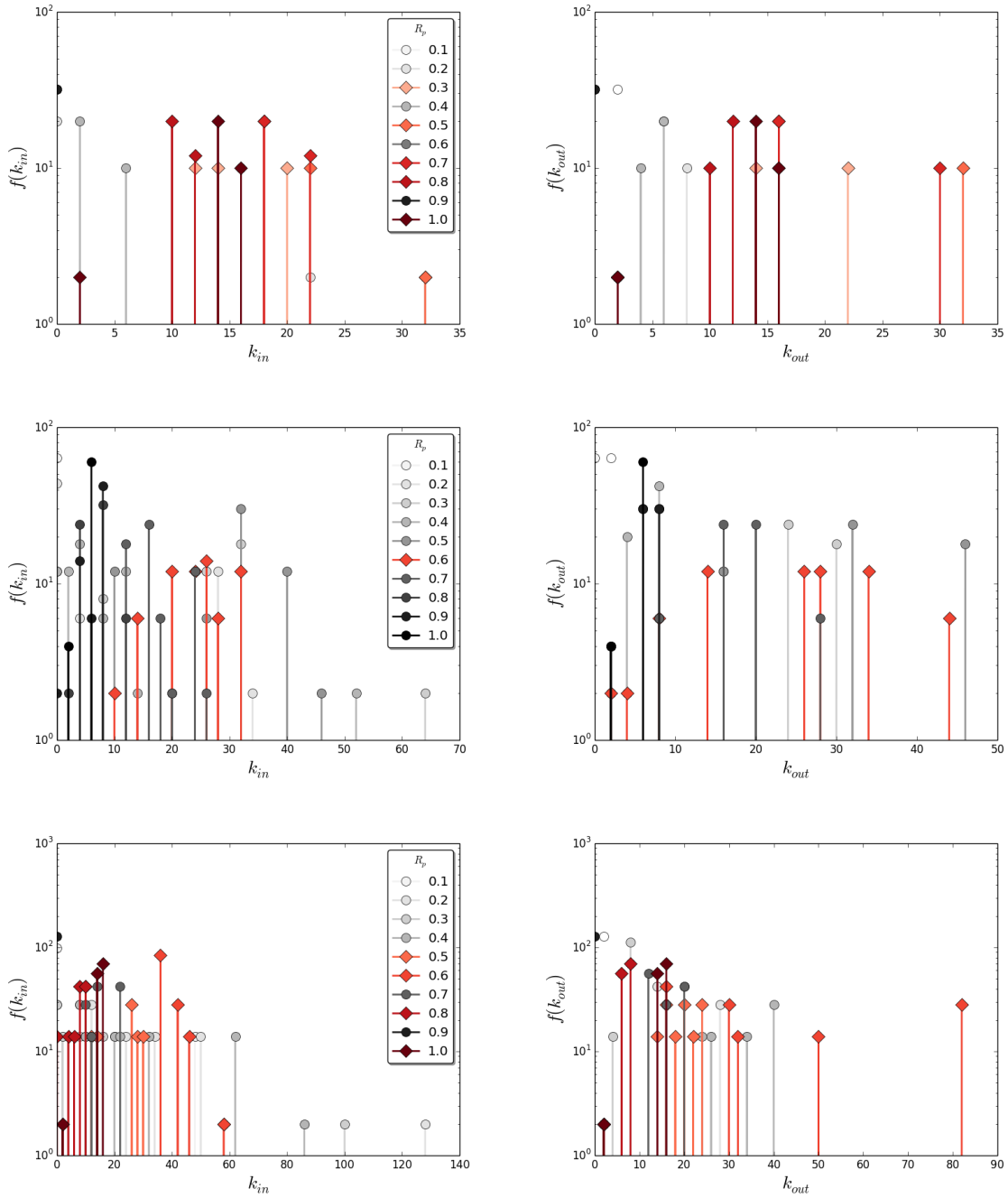


Figure 36: Frequency distribution of the in-degree (k_{in}, left) and out-degree (k_{out}, right) for state-transition diagrams for varying R_p for widths $w = 5$ (top row), $w = 6$ (middle row) and $w = 7$ (bottom row). State-transition diagrams with sampled trajectories exhibiting open-ended dynamics are shown in red, while those where no open-ended cases were confirmed are shown on a gray scale.

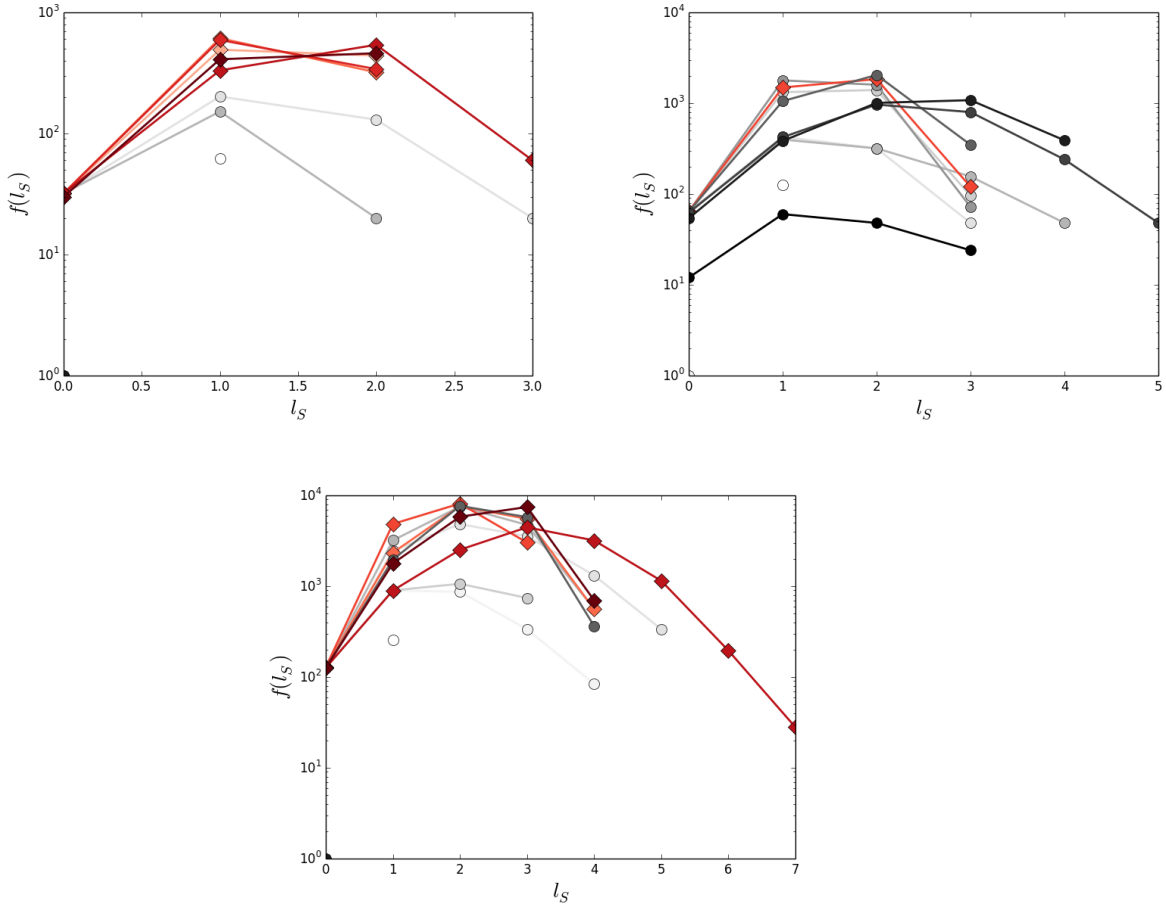


Figure 37: Shortest path (l_S) for state-transition diagrams for varying R_p for widths $w = 5$ (top left), $w = 6$ (top right) and $w = 7$ (bottom). State-transition diagrams with sampled trajectories exhibiting open-ended dynamics are shown in red, while those where no open-ended cases were confirmed are shown on a grey scale.

approximations to the principle of physical universality (or its local version) may be what is physically realized.

Our model demonstrates that state-dependent CA can exhibit local physical universality and that these systems can meet the minimal requirements of a formal definition of open-ended evolution. Local physical universality emerges as a result of the connectivity of the underlying state space. Our analysis shows for ECA rule networks, that this property is associated with heterogeneity in the in-degree and out-degree of states. It is interesting to speculate how this property relates to real living systems, which are known to exhibit highly heterogenous degree distributions across a variety of systems and scales. One pertinent example is biochemical reaction networks, which exhibit heavy tail degree distributions, with a few hubs dominating network structure (Jeong 2000). A hub in a biochemical network, such as H_2O, is hub because it can undergo many chemical transformations. It is possible heterogenous networks emerge not only because of their robustness properties as frequently hypothesized, but also due to their ability to access many configurations. This second hypothesis is consistent with evidence of an increased propensity to innovate due to the structure of gene networks and their ability to quickly traverse phenotype space with relatively few mutations (defined in terms of what reactions are possible) (**wagner:2007**; Wagner 2014).

Local physical universality is always a property of reversible rules. However the size of the subset of configuration space that has this property may be small or large depending on the particular dynamical laws (the particular CA rule). In our model, connectivity enabling large regions of configuration space to be locally physically universal can be an emergent property of the state-dependent system (as occurs for irreversible rules), which arises because of the multiplicity of the rules it can access

(*e.g.* the transformations it is programmable to perform). In particular, we equate this property with the particular informational structure of living systems. The advantage of defining local physical universality is that it permits studying systems of varying degrees of universality, and in particular to discuss cases where the ability of systems to increase their repertoire of possible transformations might change in time, as through biological evolution. One could speculate that the biosphere as a whole has trended toward increased connectivity among the possible configurations of material that make up the Earth-system (Sara Imari Walker 2015), corresponding to better and better approximations to local physical universality. This occurs through evolutionary processes that enable an increasing number of transformations to be mediated by life and its artifacts (such as technology): although the raw materials have always existed on Earth to construct satellites, it is only recently that information processing systems have evolved the capacity to transform those materials into machines and to launch them into space. The framework presented here attempts to explain some of these properties starting from the assumption that the effective laws of physics governing life are state-dependent.

4.10 Project Acknowledgments

This publication was made possible through support of a grant from Templeton World Charity Foundation and the support of the Arizona State University NASA Space Grant.

Chapter 5

DATA FROM TECHNOLOGY

Technology in today's (respectively more) globalized society has been providing us with an overabundance of data ("Big Data" 2015). More than ever, human activity can be tracked with finer and finer resolution through things like pictures and text. It is quite simple to write a script that tracks the number of Twitter users who are interested in digital humanities over time. A snapshot of a group of Twitter users is illustrated in Figure 38.

We can even track things like the changes made to George W. Bush's Wikipedia page over the course of a few years (DeDeo 2011). Because of such technological data, we are able to see large-scale biological processes in terms of innovative technological and social systems, from the evolution of patents (Chalmers et al. 2010 and Buchanan, Packard, and Bedau 2011) to the evolution of online social systems (DeDeo 2011, Oka and Ikegami 2013, and Oka, Hashimoto, and Ikegami 2015).

Technological data represents human behavior, which is arguably centered entirely on sense and experience, two very difficult things to quantify reliably. If the goal is to answer the Big Question, it's important to understand living systems across all levels of organization (microstates and macrostates, both smaller and larger). Analyzing this data could be used as a surrogate for other biological levels under appropriate abstractions. What can Facebook data tell us about the behavior of cells?

Data generated by humans is probably the best known example of open-endedness since its development is driven by human innovation and creativity (Taylor et al. 2016 and Banzhaf, Baumgaertner, and Beslon 2016). For this reason, modern data

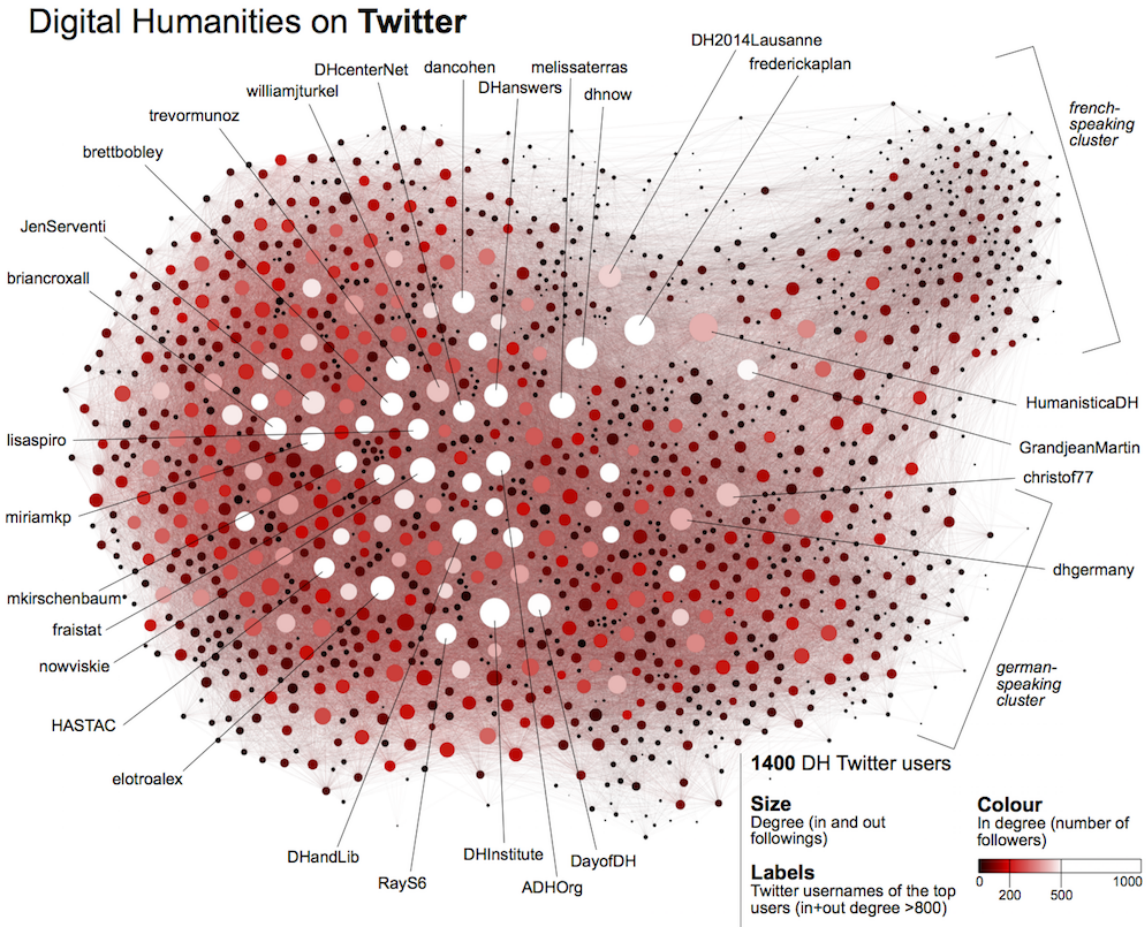


Figure 38: Twitter relationships between 1400 digital humanities researchers, engineers, and enthusiasts as of 2014. Nodes represent Twitter profiles and edges represent "following" (from Grandjean 2017).

sets make excellent case studies for living systems. Since these sets are made of living entities (humans) and elements constructed by living entities (computers, the Internet, hardware, etc.), it is one of many levels of biological organization. Perhaps understanding general properties about this higher level of organization could be useful for understanding lower levels of biology. At worst, properly analyzing the abundance of technological data could provide a highly detailed assessment of one

level of biological organization, which could have many useful implications for society and industry in particular.

To find these generalizations, some level of abstraction is necessary while also dropping "unimportant" information. In the next chapter (Chapter 6), data about a video game is abstracted into a quantifiable analysis. Many other representations of the game are equally as valid, depending on which aspects of the game are highlighted and studied. Once a system is described with some chosen variables based off relevant observables, the system can be tracked within a mapped space that is defined (even infinitely). However, in the case of Chapter 6 and many other complex systems represented by modern data, the possible states are undefined. As discussed in Chapter 1, biological states may not evolve according to a pre-defined set of states. There may be no knowing the number of possible species evolution can produce because biology might re-define its own state space as it evolves over time.

Furthermore, it is impossible to the complete state of a real-life system anyways. Knowing the state of even a video game requires one to know the state of every player and every game developer, including their mind and thoughts. This is simply not possible. It is only possible to know what data has been collected about the game over time, including the evolution of the base code. Therefore, we are once again restricted to what we are able to measure about a system by what we are able to observe. Knowledge about what players feel–such as when their in-game characters die–has been lost and yet could be a crucial part of the game's overall dynamics. Although this example is specific to one particular video game, similar problems are present in almost every set of modern technological data.

5.1 Data as Insights

In 2013, 90% of the world's current data had been created in the previous two years from uploaded videos, social media, cell phone data, and many other used technologies. At that rate, 2.5 quintillion bytes (literally a billion-billion bytes) of data were created *per day* ("Big Data" 2015). Since then, more people have been using technology and data is being created at an even faster rate. In 2016, the following estimations were made about data creation *per minute* on the worldwide internet (according to Schultz 2016):

- Over 350,000 Twitter posts
- Almost 400 hours of YouTube videos
- 2.5 million Instagram likes
- Over 3 million Facebook posts
- Almost 4 million Facebook post likes
- 4 million text messages sent

As of the writing of this dissertation, it is unclear how much this data-creation-rate has increased since 2016, but it is safe to say that is probably still increasing and is likely to continue to increase. Much of the data is unsorted, meaning it is sitting in internet-space without being organized into a scientifically-usable format. Although there is a huge surge in the amount of data that is created, hardly any of it is in an analyzable format (this is the biggest current motivation for the sudden demand of data scientists).

In attempts to skirt this massively problematic big data problem, some simply search for correlations in the data. The more data the better; no need for underlying theories since the data will speak for itself (Calude and Longo 2016). However, this is

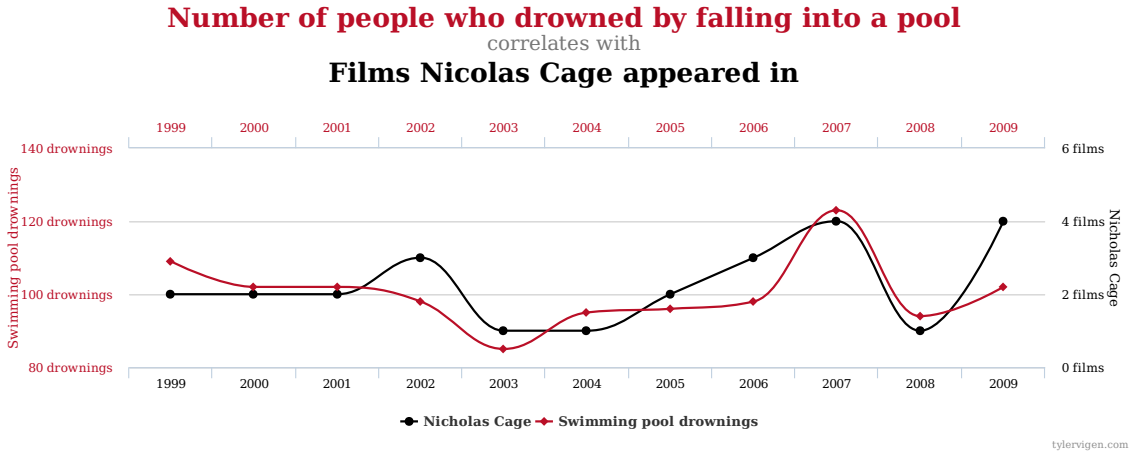


Figure 39: One example of spurious correlations from big data (from Vigen 2017).

a common pitfall in modern big data analytics. Correlation does not equal casuation and we still need underlying theories to interpret data correctly. There are several amusing graphics that illustrate where this failure can lead us, such as the correlation between the number of people drowning in a pool and the number of films Nicholas Cage appeared in (many more of these can be found at Vigen 2017).

Companies will often hire data scientists and ask for correlations in the data, particularly for useful and actionable busness insights (Hansen 2016). But without a clear understanding of the mechanisms that drive the data and how to interpret the measurements on the data (read: causation), this endeavor is likely to be fruitless. In fact, correlations can always be found in large data sets, even if the data is completely random (Calude and Longo 2016). This means that most correlations are actually meaningless. This has been known for quite some time in mathematics, in particular from ergodic theory, Ramsey theory, and algorithmic information theory (Calude and Longo 2016).

So what are we to do with so much data then? Work to find new theories that allow us to understand it, of course! Goldenfeld and Woese expect a dramatic tranformation

in the near future with this onslaught of data in Goldenfeld and Woese 2007. The way we understand the world is totally transformed when we acquire a new body of data so it is natural to assume that with this data, we can expect new theories to emerge. But this is no easy task, particularly because this transformation relies on building a theory of complex systems and biology-related-things. Yet, it may be easier to start with some homogeneous sets of data, such as data from a single video game or data from the publicly-available patent records.

Because human populations are biological, it is actually very useful to compare society-wide phenomena to traditional biological evolutionary phenomena. In particular, understanding cultural evolution is useful in terms of adaptive creativity because the same statistical tools can be used. The evolution of the patent records over time is a useful analogy to biological systems evolving under some environmental pressure (in this case, technological advances) (Skusa and Bedau 2002). By analyzing the internet dynamics of web searches and social media posts, most-used keywords (words that exclude "the" and "and" etc.) were found to act a source of information. Infrequent keywords act as a source "sink" for information, or places where information stops flowing (Oka and Ikegami 2013). Therefore, it is hypothesized that these frequent information-holding key words form the core activity for the internet, as sources of information for social media's infrequent word activity. Information flow is usually a bit easier to understand in terms of human communication because that's where our original understanding of information came from anyways (Shannon 1948). But as we will see, information is actually much more difficult to understand in different contexts.

5.1.1 League of Legends Data

One of my personal favorite data sets is from the video game League of Legends (League). This is mainly because I'm familiar with the game's dynamics and subtleties[21] but also because it is a very complete data set. Massive amounts of information about every single game played around the world over the last 6 years has been recorded and stored, ready for user consumption. In addition, data about game updates, player behavior, and server technology changes has been stored as well.

League data has previously been used to study the psychology of players (Hsu 2015), particularly on what game changes can mitiage negative player behaviour towards other players (League of Legends is notorious for its negative player community). It was found that displaying positive gameplay tips (such as "Be nice to your teammates!") in blue and negative gameplay tips (such as "Don't be a jerk to teammates!") in red found to descrease negative plater interactions. However, switching or randomizing the red and blue colors had no effect in player behavior (Hsu 2015).

The game was created and is maintained by the game company Riot Games. As of 2016, League of Legends had a staggering 100 million monthly players as reported in Tassi 2016. On top of that, it has a very large eSports presence and holds several worldwide professional tournaments and competitions. Top players from around the world join professional teams to compete for money and fame by playing League. From an evolutionary standpoint, this could be an interesting set of data to see how strategies evolve and are successful, especially at the professional level.

Collecting data from Riot Games' API servers has beome very simple and streamlined since their lastest Developer Version 3 update (Games 2017). This is as simple

[21] I do indeed play way too much League of Legends in my spare time.

as making a Python url request. After a request is made, the API returns Python-readable json-formatted data. Prior to the Developer Version 3 update, it was simpler to use Cassiopeia, a Python library to get Riot's API data into Python objects ("Cassiopeia" 2017). However, using Cassiopeia is no longer necessary. As with almost any big-data project, a relational database is necessary to organize the data into an easy-to-analyze format. Ideally, SQL is best for this job. However, Datreant is a Python-based relational database system that is useful for smaller amounts of data and is attactive for those looking to stick to Python-only solutions ("Datreant: Persistent, pythonic trees for heterogeneous data" 2017).

Riot Games is a game software developer company that strives to put players and quality games first. Their big claim-to-fame is that they strive to put players first and are very actively involved in the player community ("Riot Manifesto" 2017). Riot's developers respond to player's requests and feedback about gameplay and interact with the players on a daily basis. As a result, this adds to the evolution of the game in an interesting double-feedback way. Both the players and the game developers interact and change according to each others' behavior. A full description of a League data analysis is included in Chapter 6.

5.2 Networks for Data Analysis

One of the largest research projects specifically aimed at understanding how large social networks change over time is called the Copenhagen Networks Study (Stopczynski et al. 2014). Networks have been a very popular way to analyze and visualize a large amount of interactions between things. In social networks, nodes are often people and edges are often relationships between two people.

Once a network (also called a graph) is modeled after a real system, it can be measured in several ways. One of the more difficult aspects of network analysis is how they change over time. Many current network analyses revolve around a static snapshot of the network at some point in time and comparing it to another snapshot at a different time. While this has been pretty useful, it remains unclear how making these snapshots affects our ability to understand the underlying dynamics of the system in question. This problem only gets worse because a single system can actually be represented by several different graphs. It all depends on what is being measured, and from what perspective. So sometimes measures on graphs are actually artifacts of how a system is being represented (Zenil, Kiani, and Tegnér 2017). Many networks that represent real social systems have communities where each and every node belongs to more than one group. As a consequence, the global hierarchy of nodes does not reflect relationships within individual groups (Ahn, Bagrow, and Lehmann 2010). Regardless though, network analysis does provides a lot of insight into biological systems that would otherwise not have been realized (Rosen 1981).

There is a huge number of network measures, each invented specifically with answering a specific question in mind. Some of these include perterbation centrality, where nodes with high values of this measure were found to be crucial in group-to-group communication (Szalay and Csermely 2013), time-ordered reachability and betweenness (Kontoleon, Falzon, and Pattison 2013), and other entropy measures. Google and other web search services (but who uses the others anyways?) uses some secret modified version of page rank to sort search results (Austin 2017). Page rank can be understood in terms of how much social power a person has in their network of friends, loosely. Imagine two people, each with five different friends. The first person's friends are Barack Obama, Mark Zuckerberg, The Cubes (Mark Cuban, of course),

Will Ferrell, and B.O.B. The second person's friends have literally no other friends and don't leave their house. Therefore, even though both people have the same amount of friends, the first person has a much higher "page rank" because their friends are very well-connected.

These measures are used to probe a large variety of questions about biological networks, particularly about the difference between biological networks and non-biological ones. For examples, what does it mean for a network to be robust? Not only is this an interesting question for biology, but it is a critical question in terms of cyber security as a defense against internet attacks. An analysis of the affected nodes during attacks are only half the story. Instead, we also must consider what happens to the unaffected nodes, as stated in Tejedor et al. 2017. In fact, the dynamics of both affected and unaffected sub-networks within the whole network gives a more complete picture of what it means to be robust. In these dynamics, nodes re-rank in surprising ways, and this affects the overall dynamics of the whole network.

Regarding gene regulatory networks, which represent how genetic material and proteins interact in known ways, there are several ways to model them: network parts lists, network topology models, network control logic models, and dynamic models (Schlitt and Brazma 2007). The biological networks in one study were found to process more information than random networks (Kim, Davies, and Walker 2015). Biological networks also exhibit a scaling relation in information transferred between nodes and the most biologically distinct regime of this scaling relation is associated with a subset of nodes that regulate the dynamics and function of the networks. This subset of nodes is known as the control kernel, pictured as the red nodes in Figure 40. A control kernel nodes, when held to fixed states, evolve the rest of the network towards its most likely behavior. One interpretation is that the control kernel contains the "program"

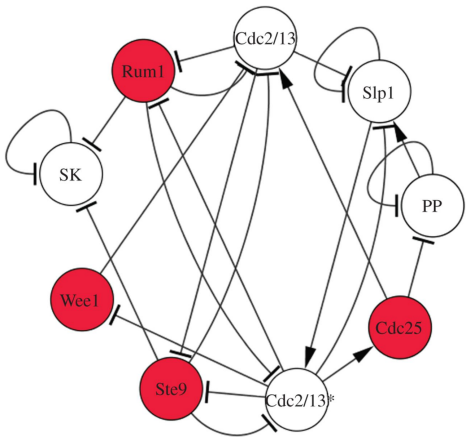


Figure 40: A network representing the protein-protein interactions in the cell-cycle regulation mechanism of *S. Pombe* fission yeast (from Kim, Davies, and Walker 2015). Red nodes indicate the control kernel.

of the whole network. Every network has a control kernel, however in randomly generated networks the control kernel is usually the entire network. Control kernels in biological networks are much smaller, which might imply a kind of "robustness". Therefore, information processing in biological networks is an emergent property of topology (causal structure) and dynamics (function) (Kim, Davies, and Walker 2015).

In the Copenhagen Networks Study of complex and densely connected social networks, fluid gatherings are organized through a stable core of individuals, which represents a social context in Sekara, Stopczynski, and Lehmann 2016. At least this way, large social network dynamics can be greatly simplified in terms of smaller social contexts. But how do these networks form in the first place? Aside from being artificial representations, they still have a good causal representation of the underlying dynamics of overall system dynamics. The evolution of complex systems may be attributed to regulation networks and niche construction in addition to path-dependent evolutionary change (Laubichler and Renn 2015). Evolving a network topology by increasing the number of connections active nodes have and decreasing the connections

that non-active nodes have could generate self-organization in natural complex systems (Bornholdt and Rohlf 2000).

Lastly, networks are also useful in understanding causality, which relates back to information and entropy from a different angle. The structure of the universe and our ability to perceive it could be actually be explainable in terms of fundamental physics (Sara Imari Walker 2015) and network theory. On a causal graph where nodes are states of the universe and are connected by possible causal transitions, some local regions have a high probability of states that contain information. These states are the most highly connected to other states, thus the underlying causal structure of reality may be a simple explanation of how living systems are able to perceive the universe and make sense of it (Sara Imari Walker 2015).

Chapter 6

REAL-WORLD OPEN-ENDED EVOLUTION: A LEAGUE OF LEGENDS ADVENTURE

6.1 Abstract

A prominent feature of life on Earth is the evolution of biological complexity: over evolutionary history the biosphere has displayed continual adaptation and innovation, giving rise to an apparent open-ended increases in complexity. The capacity for open-ended evolution has been cited as a hallmark feature of life, and also characterizes human and technological systems. Yet, the underlying drivers of open-ended evolution remain poorly understood. League of Legends (League) is an online team-based strategy game that has become immensely popular over the last six years. Because new characters (called "champions") are regularly added and the game is updated every few weeks by the game's developer Riot Games, the game never settles into an equilibrium distribution of player strategies. Innovative strategies are required for players to succeed, just as innovation is required to outcompete other organisms in open-ended biological systems. Although understanding open-endedness is crucial to understanding how living systems operate, it is often difficult or impossible to collect sufficient data to study the mechanisms driving open-ended evolution in natural systems. Online social systems, particularly games, offer ideal laboratories for studying open-ended evolutionary dynamics because of the rich data archived on statistics of users and their interactions. We focus on using data from North America's top 200 players to determine how dominance hierarchies emerge from player strategies

and how they evolve in time after an external perturbation. This is a microcosm for studying, in detail, how external and internal mechanisms can drive a real-world open-ended system. Our goal is to provide general insights that can be applied to a wide range of fields, including astrobiology and evolutionary systems[22].

6.2 Introduction

As stated throughout this dissertation, the exact mechanisms underlying open-ended evolution (OEE) are unknown. We lack a general enough model that describes all levels of biology and its open-ended and emergent properties. But since today's technology is providing us with an overabundance of data, recent work on open-ended evolution has shifted towards studying innovative technological and social systems, such as the evolution of patents (Chalmers et al. 2010 and Buchanan, Packard, and Bedau 2011) and online social systems (DeDeo 2011, Oka and Ikegami 2013, and Oka, Hashimoto, and Ikegami 2015). Online video games have rich dynamics as well, with high-resolution data available on player statistics and strategies.

In this chapter, data of the best North American players in the online video game League of Legends (League) is analyzed as a tractable model for studying a real-world open-ended system. League is an online community of players where millions of players play with and against each other on a virtual battlefield, pitting 140+ in-game characters against one-another in team battles. The game is developed by Riot Games and is regularly modified (through the game's base code) every week or two. The interactions between players within the game and external interventions by Riot Games lead to a dynamical system where player strategies never achieve a fixed, stable

[22]This chapter is adapted from it its original publication Adams and Walker 2017)

state. Because of this, the entire system of League (Riot Games + players + League software + hardware) has many 'life-like' properties that resemble features seen in other biological systems.

As a system, League updates itself according to its own history and its current state, making it a self-referential system. Memory of the past is somehow stored in the game's current state. The system of League can most simply be partitioned into two interacting subsystems: the players and the game developers. Both these subsystems influence each others behavior much like an organism (game) and its environment (developer) interact in a highly non-linear way. The game as a whole does not evolve in time according to a pre-defined state space. Since the game developers are not limited in the number of changes they can make to the game, the game evolves in an open-ended way.

6.3 League of Legends is Open-Ended

League of Legends is an online team-based strategy game that has become popular over the last seven years and is currently the most popular game of its type (Games 2016a). In every instance of a single match, ten players are randomly partitioned into two teams of five, the "Blue" and "Red" team. The two teams battle against each other on a virtual battlefield, shown in Figure 41. Each player picks a single in-game character ("champion") to play on the map throughout the entire match. Three examples of champions are shown in Figure 42. Players use their chosen champions to compete against the enemy team for resources on the map.

Two instances of the same champion are not allowed in a single match. Resources

Figure 41: League of Legends' most popular virtual battlefield, Summoner's Rift. Team bases are located on the opposite corners of the map. Player champions spawn at their respective teams' locations. There are three main spatial "lanes" where players generally battle: Top, Mid, and Bottom lane. For scale, the champion "Teemo" is shown in Mid lane.

Figure 42: Three of the 140+ example champions players can choose from: "Talon", "Teemo", and "Thresh".

are used to upgrade champions to level them up throughout the duration of the match. These are collected around the map and can be earned by killing enemy champions or other objectives on the map. The main goal of the game is for one team to destroy the other team's base. This requires a high level of cooperation and trust within individual teams of players. The more matches a single player wins over time, the higher the player climbs up in skill ranking.

Riot Games, the game's developer, updates the game every 1-3 weeks. They add new champions, update old champions, change properties of the map, and change other aspects of the game via code updates. Most of these changes are aimed at making the game more fun for the players ("Riot Manifesto" 2017); listening to player needs/wants to make the best accommodations possible. Thus the majority of Riot Games' changes are based on feedback from the players, and reflect the current state of the game.

Figure 45 shows a schematic of the players explorating possible game as constrained by the game's code, changed by Riot Games. The possible number of game strategies is undefined because it depends on the number of possible changes Riot Games can implement in the game code. The game's code depends on several unaccounted factors such as technology that is implemented, creativity, and even the economy. In this sense, the game as a whole (players, platform and Riot Games) is open-ended, since its evolution will never repeat (Banzhaf, Baumgaertner, and Beslon 2016, Ruiz-Mirazo, Peretó, and Moreno 2004, and A. Adams et al. 2017).

However, in a single implementation of the game's code, the players have a finite number of strategies to choose from in this undefined state space. By defining a constant, unchanging constraint on the strategy space, the dynamics are guaranteed to repeat and open-ended evolution is not possible due to the Poincaré recurrence

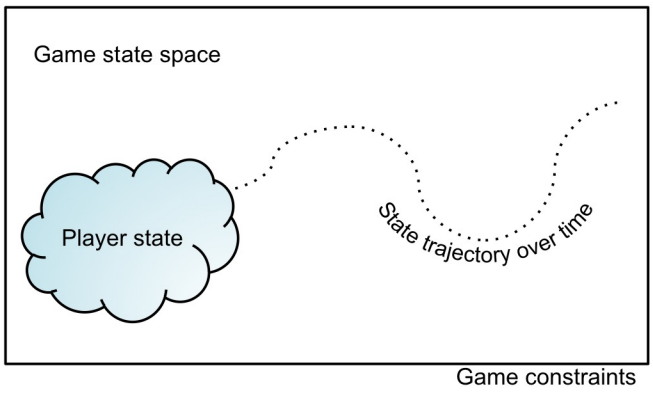


Figure 43: A cartoon representation of how players explore the possible strategy state space over time. The game state space represents all the possible strategies players can choose from, constrained by the current game's code. A current player state represents what distribution of strategies are being used at some given time.

theorem (A. Adams et al. 2017). Changing these constraints often and drastically on a regular basis guarantees that player strategies do not repeat, thus the player dynamics are open-ended as long as the constraints are changing (A. Adams et al. 2017).

In Figure 44, the state-analogy is illustrated along with the double-feedback mechanism between players and Riot Games. As players explore strategies within the game's constraints, Riot Games "measures" the strategies by making an observation and changing the game's code as a result. This affects the players by changing the number of strategies available to them. It is unclear how a signal is translated into an

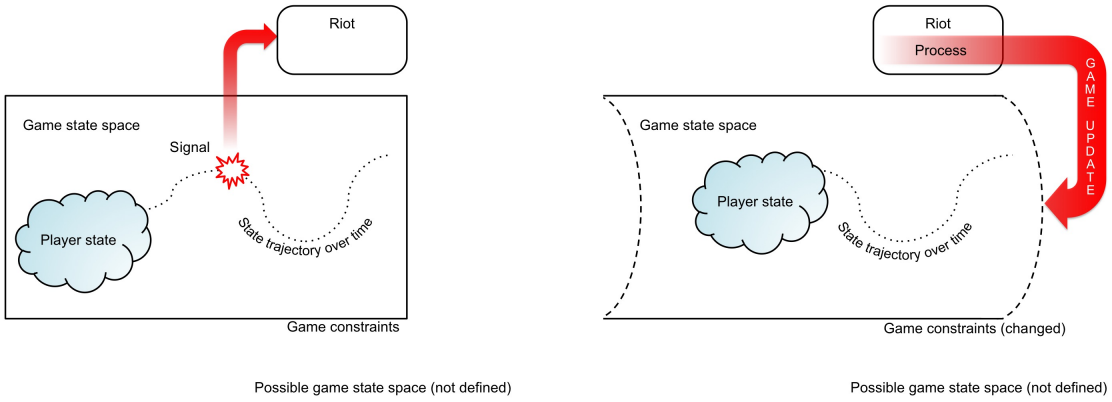


Figure 44: **Left:** At some point in time, Riot Games measures the current distribution of player strategies. **Right:** Riot Games responds by updating the game's code, thus changing the space of possible strategies players can use.

act on the game's current constraint, and Riot Games might not use a single function to determine it. This analogy makes it easy to compare it League to biological systems.

6.3.1 Collecting Relevant Data

League is a very complicated and layered game in its own right. There are three maps, seven match modes, and eight tiers of player skill. For this analysis, we only collected data from players that have the highest skill ranking on the North American server. League's skill ranking system is crowned with the Challenger tier: Only the top 200 players on each server occupy this skill tier. If a player is in the Challenger tier, they are most likely a professional player and stream their matches on Twitch (an online streaming service) for money. Several tournament players are in Challenger tier as well.

In League's player community, Challenger players often set gameplay trends. Since they are the best-of-the-best players, lower-level players often imitate their strategies

and gameplay styles in hopes of climbing the skill ranks themselves. Hence, we assume that Challenger players are the players' main strategy trend-setters. That there are other aspects of lower-skill gameplay that are not addressed in this chapter. The main point of this study is not to determine if this system evolves in an open-ended way, nor does it explore what drives the open-endedness. Rather, the point is to determine how this system evolves mechanistically and if these mechanisms can be generalized to other levels of biological organization.

We also only consider matches that were played on League's most popular map, Summoner's Rift. Since we want to ensure these players are playing their best (because players can purposefully play bad games to entertain an audience), we only consider Ranked matches. Ranked matches determine if a player stays in Challenger tier or is demoted to a lower skill tier and are therefore taken more seriously by the players.

Here, we consider two consecutive game "patches", or when the game code is changed by Riot Games. Between the times that the game is patched, the game's constraints remain unchanged and constant. As a starting point, it will be easier to assess two separate time periods in the data where the game constraints remain unperturbed by developer intervention. These patches, named 7.2 and 7.3 by Riot Games, were used 6 Jan 2017 to 7 Feb 2017 and 9 Feb 2017 to 22 Feb 2017, respectively. One day between the two patches was discarded since the patch 7.3 was deployed that day.

6.3.2 Methods

All game data is made freely available to the public by Riot Games, and can be downloaded thorough Riot Games' server API (Games 2016b). To save time,

we used a pre-written Python 3.X package called Cassiopeia ("Cassiopeia" 2017) as an interface to connect to the API. Datreant ("Datreant: Persistent, pythonic trees for heterogeneous data" 2017) was also employed for fast and easy data sorting and exploration.

6.3.3 Internal Mechanisms

Challenger tier players generate data by playing matches (ranked matches on Summoner's Rift). Between patches, the game is fixed in the sense that it has a static code underlying the game and a pre-defined state space of strategies; there are only so many different things players are able to do with the game's current code. In physics, systems typically evolve under a pre-defined state space. This makes them easier to understand conceptually. Equations are used to quantify how systems move through and explore the pre-defined state space. For this reason, this part of the data is the best place to start with the intention of understanding the game's dynamics as a whole. In other words, we begin our analysis by using player-generated data between Riot Games' game changes, where the game's code is unchanged and the constraints that the players experience are constant.

How are the players moving through the strategy space? Presumably, players want to win matches and are picking strategies to help them achieve that goal. So is the players' trajectory their way of discovering optimal strategies to win the game? There is a subtle conceptual pitfall here based on what the data represents. Despite Riot Games' attempts to make all of its champions equally competitive, there is no reason to assume there is a pre-existing optimal champion or strategy to use. All knowledge about how players explore strategies (including Riot Games' knowledge) is

based solely on player data. There are only a finite number of possible states the data can be in at any given time. There is also no explicit "best" strategy that is optimal at any given time. However, players supposedly seek a type of optimum by exploring possible strategies and using the ones that win the most games. This could be an analogy for biology since organisms and species evolve by exploring the possibility space of phenotypes and use the ones that are the most "successful" at a particular time.

There are several different ways to quantify how the players explore possible strategies and evolve the current "metagame." The term "metagame" is akin to the set of most successful strategies being used at a given time. In an attempt to quantify the metagame, *which champion beats what other champions* will be the main definition of a successful strategy.

Champion selection is a core mechanic of selecting a strategy. If a player sees the enemy team picks the champion "Teemo", the player is likely going to pick a champion that can beat "Teemo," such as "Talon." This decision is guided by knowledge from past experiences. In previous games, the player was successful in beating "Teemo" with "Talon" (or perhaps participated in games where they observed this interaction between other players) several times. In other words, players have some general knowledge about the current metagame and how to leverage it towards their advantage.

Since there is no explicit optimal strategy, only player-generated data is available to determine which champion beats who. Let's say champions A, B, C, D, and E are used on Blue team, and champions V, W, X, Y, and Z are used on the Red team for a single match. Blue team wins the match. We say that in this match, champion A beat all Red team's champions (V, W, X, Y, and Z), W beat all Blue team's champions (A, B, C, D, and E), etc. This is a generalization of high-level play, since all players

Figure 45: The champion “Teemo” beat the enemy champion “Tristana”. On a graph, this is represented by an arrow going from “Teemo” to “Tristana”.

on Red team interact with all players on Blue team frequently during a single match. This also simplifies the analysis.

This can be represented as a network, where if the character Teemo beat the champion Tristana, the interaction is represented by a directed edge from node “Teemo” into node “Tristana”, as shown in Figure 3.

For an aggregate of matches, edges have weights to represent exactly how badly “Teemo” beats “Tristana” during those matches. These weights are quantified in the next section.

6.3.4 Meta Strategy Trends

When League was first released in 2009, any champion was played on any spatial location on the map by the players. As time passed, players decided that having certain types of champions on certain locations of the map was preferred. Since the community made this decision, a team’s general spatial organization has not changed

for over 6 years. In some sense, players reached a consensus on a firm "meta", the player-community term for a persistent and uniquely identifiable strategy trend.

Within this firm meta, metagames are regularly explored. For example, high-damage characters might be very popular during one patch. After some time, the champions that easily beat the sub-meta champions (tough champions that do not take damage) become popular, simply because they counter the current metagame. Thus, the counter metagame becomes the new metagame. This minor metagame rotation is reminiscent of negative frequency-dependent selection in biology (Allen and Clarke 1984). An agent-based model could be used to further explore these metagame dynamics as emergent strategy dynamics among players.

Since players respond to current metagame by finding counter-strategies within the firm meta, the system retains a type of memory. There is an underlying mechanism that causes new metagames to emerge while allowing the firm meta to persist, even if a better firm meta may exist.

6.4 Constructing Networks

To represent the players' internal dynamics under constant constraints (during a single patch, when the games' code is unchanged), a champion counter network is created to see how champion "power" changes over time. In this context, power is used to describe how dominant a champion is over other champions, given a set of match data. This is a simple way to represent player strategies. Dominant or powerful champions are often more likely able to win games, regardless of the rest of the team's performance. Players often consider these champions to be powerful,

while other characters could be considered under-powered ("Who would ever play *that* champion?" a player might remark).

At any given minute throughout the day, anywhere from 10 to 100 matches are being played by challenger-tier players. Matches can last anywhere from 20 to 60 minutes (a hard minimum of 20 and a rough higher estimate on the 60). It is unclear if the game has a natural time scale, which makes it difficult to identify individual time steps. All matches were binned into days according to their start times. This is arbitrary, but since Riot Games' changes occur anywhere between 1 and 3 weeks, it seems like a reasonable time unit for network analysis.

Edge weights between champions i and j are constructed for a given day:

$w_{i \to j} = popularity \times winrate$

$w_{i \to j} = \frac{matcheswithivsj}{totalmatches} \times \frac{numberofmatcheswhereibeatj}{matcheswithivsj} = \frac{N_{i \to j}}{M}$

Since edge weights change on a daily basis, the network is dynamic with respect to time.

6.4.1 Power

For this analysis, eigenvector centrality (EVC) is used as a representation of a node's power. For a weighted adjacency matrix W with EVC values xi, eigenvector centrality is defined as:

$Wx = \lambda x$

$x = \frac{1}{\lambda} Wx$

$x_i = EVC_i = \frac{1}{\sum} nj = 1 w_{i \to j} x_j$

Google uses a modified version of EVC to determine the webpage rank as the result of a Google search (Austin 2017). Colloquially, EVC represents how much influence a

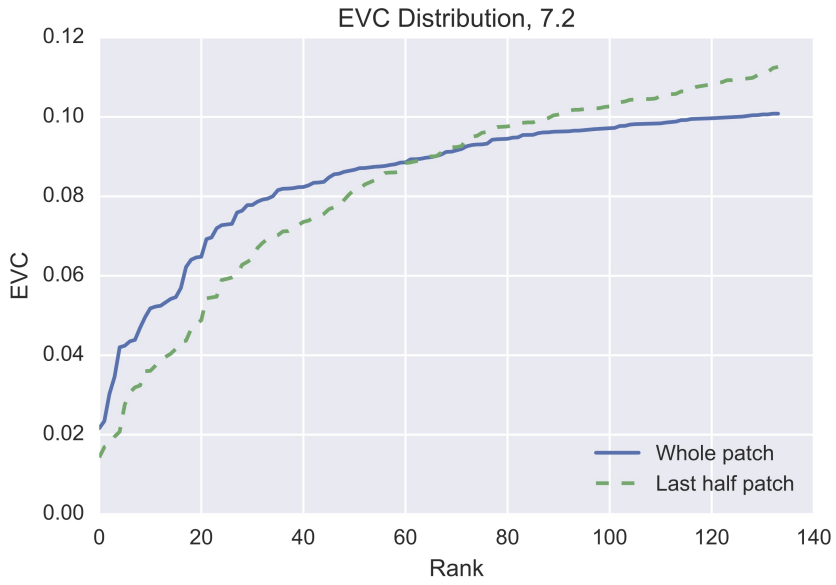


Figure 46: The rank-ordered EVC distributions of champion nodes over patch 7.2. The blue line is the distribution over the entire patch and the green dashed line is the distribution over the last half of the patch.

node has in a network, as described earlier. It not only considers how many in-degrees or out-degrees a node has, but the degree of nodes it is connected to as well. A node with 5 connections has a higher EVC value when the nodes it is connected to are also highly connected. With this metric, we are interested in two questions to understand how a champion power hierarchy is formed and maintained. (1) What happens to the power over time in champions that were changed in the last patch? (2) What happens to the power over time in champions that are going to be changed during the next patch?

To gain a general sense of the distribution of power in the game, all data for a single patch was aggregated into a single network. The distribution in Figure 46 shows the EVC ranking of champions over patch 7.2. This represents a larger time scale than individual days.

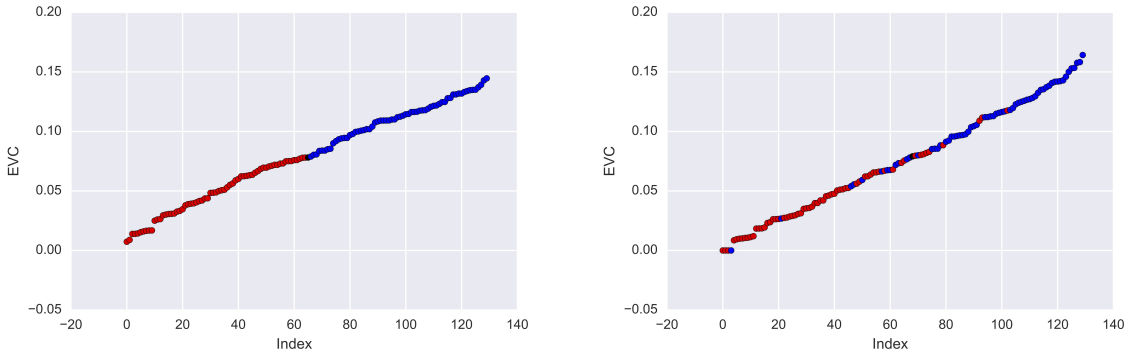


Figure 47: **Left:** The rank-ordered EVC distribution of champion nodes for the first day of patch 2.7. Points on the upper and lower half of the distribution are colored blue and red respectively. Colors simply correspond to their rank position on the first day of the patch as the top 50% and bottom 50%. **Left:** Without changing the colors of the nodes, the distribution of EVC is shown for the last day of the patch.

The same analysis was completed for individual days, but each individual day had an approximately linear distribution. Since the distribution over the whole patch is non-linear, this indicates that points on the linear daily distribution move up and down the distribution day by day (Figure 47). This figure shows the same ranking of EVC values, but on the time scale of individual days. This distribution is linear, as opposed to the non-linear distribution for the larger time scale.

The majority of champions on the upper half of the distribution continue to remain on the upper half from day to day. The non-linear distribution on a larger time scale and the linear distribution on a smaller time scale remained largely invariant for their respective time scales. However, the non-linear distribution from the larger time scale suggests there may be an underlying dynamic on how the linear daily distributions change point ranking. This phenomenon may indicate a type of dynamic equilibrium in the game's dynamics, which is characterized on a larger time scale by a non-linear distribution. The EVC values for individual nodes over time were analyzed

to address this phenomenon. No clear trends were found for the accumulation of nodes (champions), so it was more fruitful to analyze specific champions, particularly ones that were changed during the two patches (Figure 48).

From Figure 48, it is unclear whether Riot Games' patches were beneficial to the "power-balance" of the corresponding champions. In other words, it is unclear whether Riot Games' patches accomplished the goal of adjusting the champions towards a more equal distribution of power. In Figure 48a, "Akali" was buffed (made better in some way) and "Graves" was nerfed (made weaker). Their EVC values after the patch 7.2 changes seem to suggest the opposite. Only a few champions seemed to have benefited from their changes. Figure 49 suggests a better indication of what champions need future changes for the majority of champions, though these changes still missed the mark on champions like "Rengar" and "LeBlanc" since their EVC values decrease over time, yet they are nerfed anyways.

Since champion power levels depend on each other in a very intricate way, buffing or nerfing a particular champion doesn't seem to have the desired effect on that given champion. For example, if champion A beats champion B, nerfing champion A is an indirect buff to champion B. Since this system is highly connected, making changes to the game can perturb the network in unforeseen ways. A deeper understanding of how elements of the system (such as champions in this context) affect each other rather than analyzing the element dynamics themselves seems to be a more insightful direction of inquiry than a network approach. Although networks are useful in understand the direct relationships between entities, an agent-based-modeling approach might be more useful in understanding how these entities interact in a complex system.

For example, we could model agents who randomly select champions based on a

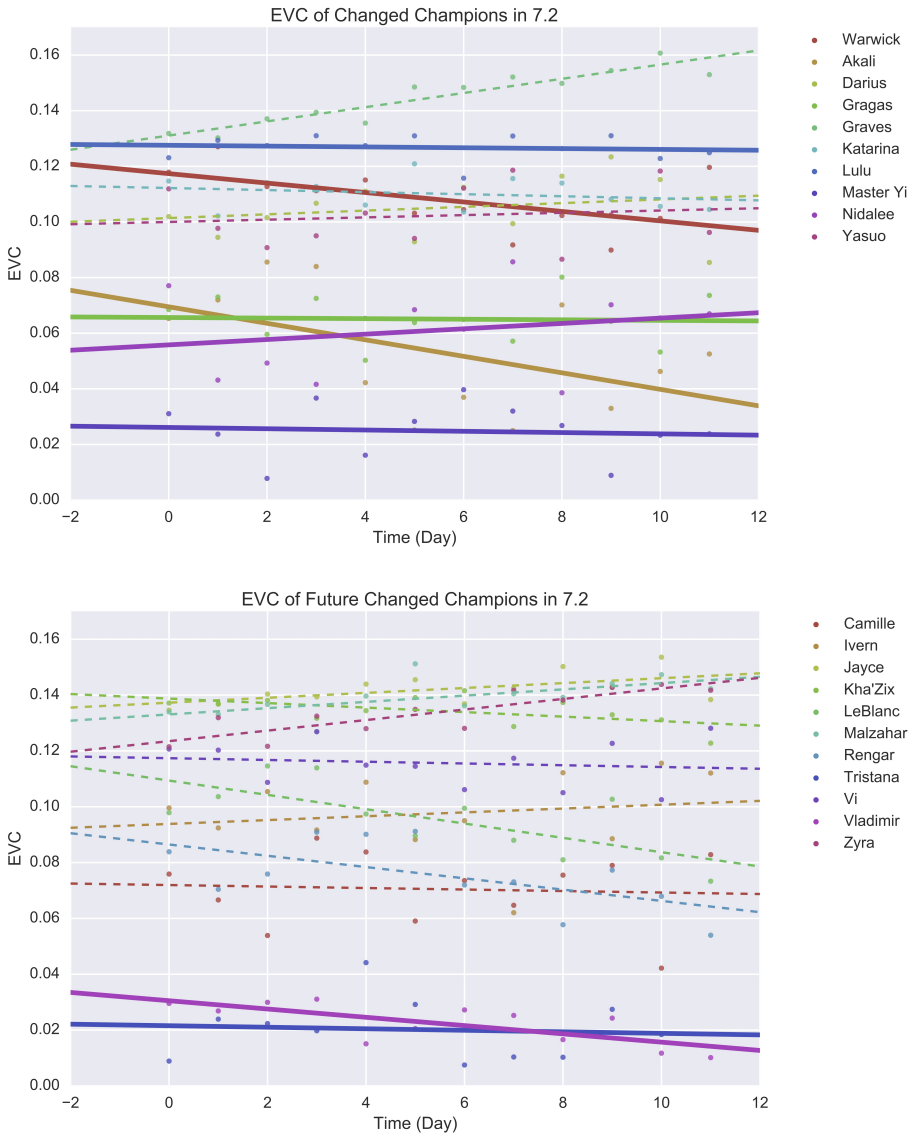


Figure 48: The EVC values for select champions (colored lines) over time. Dotted lines indicate champions that were "nerfed" (made weaker in some way) and solid lines indicate champions that were "buffed" (make stronger in some way). The top panel shows the champions that were buffed or nerfed at the deployment of patch 7.2. The bottom panel shows the champions that were buffed or nerfed at the deployment of patch 7.3.

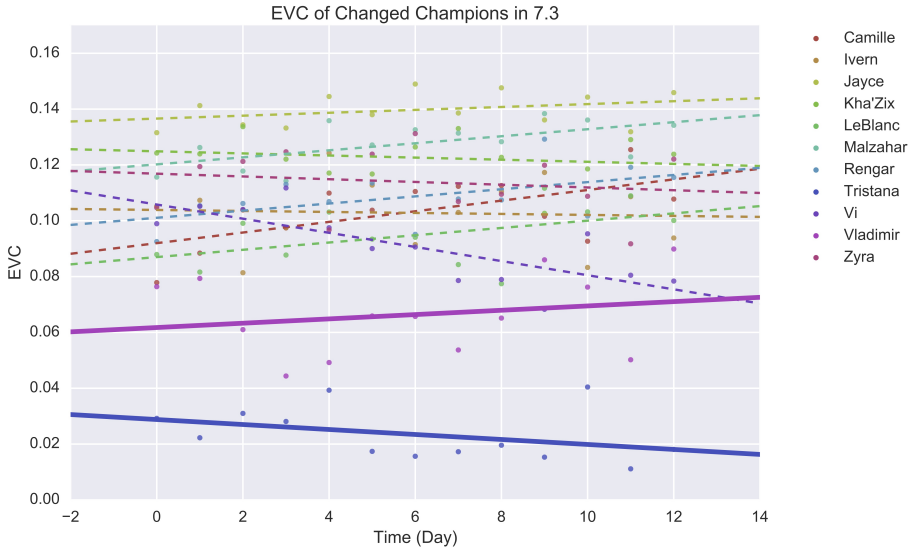


Figure 49: The EVC values for select champions (colored lines) over time. Dotted lines indicate champions that were "nerfed" (made weaker in some way) and solid lines indicate champions that were "buffed" (make stronger in some way). These champions were buffed or nerfed at the deployment of patch 7.3.

distribution derived from the data. Match outcomes could be decided on factors such as the agent's skill and data-based outcomes on which characters win against which other characters. From such a model, are the same EVC distributions produced over the rank of champions? We could also incorporate feedback, updating the champion-selection distribution based on the resulting EVC distribution. Such a model will be explored in the near future.

6.4.2 Player Perspectives

Are players aware of the entire network's structure at any given time? Players only observe interactions between 10 nodes per match, so it is unlikely they observe interactions that include all the 140+ champions every day. Players observe a smaller network from only the matches they participated in. If their goal as a player is to

win, are they increasing the EVC value on their champions' nodes in their smaller network of observed interactions? How do the EVC values of these smaller networks compare to the EVC values of the whole network?

In addition to these questions, it's important to highlight the two different kinds of League players. There are players who mainly stick to playing a single champion, and players who play whatever they think will help them win matches. We denote these players as Type 1 and Type 2, respectively. Type 1 players who stick to a single champion over time have more influence on a node. They actively maintain the state of a node and its connecting edges. In some sense, this is a lot like a system's stability. Type 2 players who only play certain champions to win and the player who are more exploratory and perhaps even more innovative. Their presence contributes to nodes that are likely more powerful. In another sense, they're a system's innovation-apparatus.

6.5 Wrap-Up

So far, we've only considered a few components of League's dynamics. Instead of using EVC as a metric, it could be fruitful to explore perturbation centrality measures (as in Szalay and Csermely 2013), since game changes occur frequently. Our next immediate goal is to explore other ways of quantifying the players' dynamics besides using a network, such as the agent-based models discussed previously. In addition, the idea of smaller player-observed networks described in the last section can be studied for practical purposes. Do players have unique "personal network" unique to their own match history or is the network topology similar for any given Type 1 player? It could be possible that if players have personal networks, cheaters can be detected.

The most common form of cheating is when players purchase high skill-level accounts to use them as their own. Another is using external code to artificially improve a player's skill. Cheating could be flagged by tracking sudden topological changes in a player's personal network.

Understanding the rest of Figure 2 is beyond the scope of this paper, but can be addressed in future studies. In particular, what signal is Riot Games gathering from the players' dynamics? How is this signal processed and translated into game updates? How do the game's constraints on the players evolve? Finally, how do all these dynamics interact and aggregate to evolve the entire game as a whole? One could argue this data could be a useful compared to population dynamics data—both measured from real systems and generated from computer models. In League however, this data exists on the boundary between real and computer-generated: real human players are using computers to interact with each other through software and the Internet. The game encompasses both types of data used in population dynamics studies. Returning to the negative-frequency dependence selection analogy to describe player strategy dynamics, it would be useful to see how League data compares to both real data and purely computer-generated data.

Physical laws as we currently understand them are insufficient to describe biological phenomenon such as heredity, adaptability, and the number of global tweets per minute. If we were to somehow rewind the universe to the point of the Big Bang, is the presence of life somewhere in the universe inevitable? Perhaps the reason this quest is so difficult is our lack of a fundamental understanding of biological data, although this is certainly up for speculation.

6.6 Project Acknowledgments

Special thanks to Riot Games' API team for allowing all League data to be public. Here's a shout-out to David Dotson, the creator of Datreant. This project would not have been possible without his help. We would also like to thank the team at Cassiopeia for their helpfulness. A.A. and S.I.W. were supported by a grant from the Templeton World Charity Foundation.

Chapter 7

FINAL REMARKS

So what is to gain from this dissertation? Mainly, it is not simple to analyze data that represents complex and/or living systems. This is largely because data is merely an abstract *representation* of a collection of real things, which is perhaps the reason why its analysis has not yet yielded to a comprehensive theory of living systems. Of course, this could be due to the fact that we simply need to do more data analysis– a *lot* more– although this approach comes with its own set of problems as well (Calude and Longo 2016 and Buiatti and Longo 2013). In short, given multiple large sets of numbers (multiple sets of data), correlations are always guaranteed to appear (Calude and Longo 2016), even if they are, for all practical purposes, absurd (Vigen 2017).

In this concluding chapter, I explore a new approach to representing the world so that better data (more encompassing data for real living systems) can be collected. This is by no means rigidly mathematical, nor founded on solid theory. Instead, this is simply an exercise in simplifying what has been discussed so far in a way that is deliverable in less words and could serve as a starting point for future, more rigid mathematical discussions on living systems.

There seems to be a general agreement that living systems need at least three components to exist: (1) A physical substrate to exist in, like molecules and atoms, (2) a way to process information in some way through sensing, (3) and a meta-rule that governs how information processing should change over time, e.g. evolution. However, the exact relationship between these three ideas remains largely unclear, particularly how information processing and its meta-rule emerges from matter. Are the laws of

nature somehow embedded within current mathematical models? This is likely not the case. Consider the quantum description of a hydrogen atom. It is highly unlikely that using the same description for the Krebs cycle, if even at all attainable, would be of any use for understanding the aerobic processes of a cell.

7.0.1 Physical Substrates

Coulomb's law does not exist where there are no charges[23]. Therefore it may be impossible to find the laws of life in systems that are not living. A single bacteria in a petri dish may not actually be a living system even though it is, in fact, alive. There should be a strict distinction in our awareness of "life" and "alive". Although a single molecule may not be considered alive, such as RNA, it is considered by most be an essential part of life. The current search for extraterrestrial life is focused mainly on the detection of biosignatures, which may not be alive themselves but are an indication that life is present. Without life, those biosignatures would not be present, whatever they may be. A mahogany table, which is also not considered alive, would be considered a biosignature of life since it is unlikely to be found where life is not present. Yet, biosignatures are significant to the environment of living systems (think of your own home). On a cellular level, if we are prepared to say an RNA is life although itself is *not* alive, then we must be prepared to accept entities produced and used by life as *a part* of life. That is to say, a mahogany table is certainly a part of a living system since it has been generated by life and is used by life.

It is therefore too difficult to delineate the boundary between life and its envi-

[23]This is a deep, deep philosophical argument, but here I am strictly asserting a claim on this side of the argument.

ronment. This is my first preposition: that we should move away from drawing a boundary around life and its environment. There should only be physical pieces of reality, whatever those may be, whether or not we can sense them.

7.0.2 Information

For humans, reality is defined by what we sense. Even though there is a strong assertion to announce that objective reality exists, we can only attribute this to the agreement of common observations between people. This consistency, although useful for constructing theories about reality, is not reality itself because it only exists within observers.

Sense is the capturing of information. Information, in some way, can be colloquially understood as "a difference that makes a difference" (Brender 2012). As far as this can be understood using the theories of computation and information, this includes necessary things like memory, state, and some sort of interactions between entities. Living systems have this ability to sense and to perceive information about internal or external things. Being able to collect this information greatly depends on the observer's ability to sense and perceive, which could be attributed to the current state of the observer.

7.0.2.1 Memory

Within physical states (or configurations/arrangements of matter), the notion of memory is embedded. In computational theory, memory is the preservation of past states through the current configuration of bits in a current state. A system

has memory if it retains properties about its previous configurations in its current configuration. However, the movement from bits to the real physical world is difficult, since it moves from an abstract realm to one laden with specific details. In the physical world, bits are understood by human notions of yes or no, up or down, left or right. If I lay the popcorn bag this-side-down then it is distinguishable from laying the popcorn bag this-side-up. The world outside computers preserves our understanding of memory in the way that if a person changes something, then that change can be seen for some amount of time. If I change the facing of my popcorn bag, then that action is transcribed in the world by the bag's facing[24]. In this sense, a person's memory is no longer limited to their own brain. A human's physical interactions with the world ensures that memory is preserved in the world in some way.

As living systems change external properties about the world (the arrangement of things within a room, for example), cause-and-effect are greatly entwined with the idea of recovering the past from physically inscribed memory. Imagine a banana taped to a ceiling of a kitchen. How did it get there? Clearly, there must be some cause for the effect to having a banana affixed to the ceiling. The number of causes is also relatively large, even on a macroscopic scale; *who* put it where, rather than what arrangement of individual atoms put it there. This loss of information relevant to an individual observer has been postulated as the origin for the arrow of time (Maccone 2009). This phenomenon is most commonly studied in thermodynamics and statistical mechanics: Given a current state $s \in S$, where we can assume S is finite, albeit unimaginably large if the universe is finite (otherwise we assume S is infinite), the number of possible causes $|C|$ from the point of view of an observer $o \in O$ is

[24]This is also the purpose of keeping a notebook. By changing the properties of the physical world (writing in a notebook), the mind's thoughts are stored elsewhere than the mind itself. A person can access the notebook (physical memory) instead of preserving the thoughts within the mind.

greater than 1.

$$|C|_o > 1 \quad | \quad s \tag{7.1}$$

7.0.2.2 Perception

Complexity theory is so far constrained to our own perception and measurements of a system that we ourselves are embedded in. Knowledge can only be constructed by our observations, which are constrained by our physical senses. As first-order observers, we may not be in a position to provide an observation of complex systems that has enough information to construct a consistent theory. If only we could become second-order observers and gain a bird's-eye view of a complex system then we might be able to gain more information to explain away what is currently unexplainable.

Imagine an event that occurs within view of multiple observers. Do all observers perceive the same things? Unless each one of those observers is in exactly the same physical state (location and molecular configuration) then the answer is no. Yet we, as humans, assert there is consistency between us in various ways. The sky is blue, concrete hurts when fallen upon, and Nickleback is a terrible band. Even though we all receive different information about our world, we are able to use language that excludes minute differences to come to an agreement in perception.

This is the strength of the scientific method. It provides a set of requirements that multiple observations must meet in order to make a statement about the world. This is the basis that theories emerge. Newton's First Law is the language used to describe many many observations made under similar constraints. It is under this assumption that the word "theory" can be used to describe an agreement among observers about an event, or set of events, in a general sense. Within this context, observations refer

to the ability to perceive information about something, internally or externally to an observer, which depends on the current state of the observer and the ability for the observer to perceive. In other words, perception can be an attribute of the state of an observer, which is a physical entity.

7.0.3 Observers

The observer is a physical entity made of physical components. The fact that it can "make observations" is attributed to the fact that it can be physically manipulated by external factors, much like a hard drive can perform read/write functions given a command. In this example, a hard drive is certainly an observer, since it perceives commands from a computer's user to change some aspect about itself. An observer could therefore be defined as a physical entity that undergoes differences when it receives a signal that prompts it to do so. However, this depends entirely on the signal as it does the current state of the observer. More on this later.

However, this implies that there is no differentiation between data and programs, or states and laws. And why should there be? Although it would be a positivist approach to assert that the laws of a system are the state of the system itself, this is not unlike many current theories in science. Coulomb's law does not exist where there are no charges, after all. Although this approach is insufficient to explain phenomena about the teleological nature of living organisms, it is at least consistent to appropriate the laws of nature as observed properties of nature itself. One approach in biology is that the structure of a molecule is most likely its function (Abbot 1916); if a protein looks like a turbine, it most likely is a turbine. For macromolecules and smaller (at the very least), their physical attributes are enough to explain how they will interact with

other entities of the exact same type. Even of interactions with entities of different types, such as ribosomes interacting with cholesterols, the rules of their interaction depends entirely on both entities' physical states without external perturbations. In the absence of external forces, both entities dictate how each entity will evolve to a future state.

However, it is completely absurd to discuss entities in vacuums, devoid of any interactions and messy, multi-leveled perturbations with the goal of understanding life. Rather, the world is full of observers on many levels of organization where elements of one observer may also be an element of another observer on the same or different scale. Thus the boundary around observers is not well-defined. For this reason, it is even more difficult to bound an observer from an environment. Instead, we must either retreat from the notion of observer, or understand the observer in terms of an emergent sense. In the latter case, it may in fact be true that observers have blurry or non-existent boundaries on many levels. For the former case, it is difficult to justify a world without observers. Although this description of the world may be useful, we as humans consider ourselves to be observers and it would be too difficult to traverse a path towards that agreement.

7.0.4 Emergence

To summarize the discussion so far:

- Living systems are composed of physical things and entities,
- Memory, the ability to sense/perceive, and information are attributes of physical configurations (states) of physical matter,
- Observers are composed of physical substrates and have the ability to sense, and

- The boundary defining an observers is not well-defined on many levels of scale and organization, along with the boundary between an observer and its environment, and

Under the existence of physical entities, information can be understood as the physical configurations of those entities and how those entities interact. Yet, we are still missing the third component of living systems: How do information-processing rules change over time? In other words, how does emergence in the most general sense (emergence of levels of organization, functions, processes, etc) happen in a physical system?

Since we are no longer assuming any explicit boundaries between entities and their environments, as well as boundaries between multiple observers, we are left with entities themselves. For completeness, all entities should be defined as sets of entities:

$$e = \{e\} \tag{7.2}$$

and all the states of entities are laws that dictate their evolution:

$$s_e = f_e. \tag{7.3}$$

The interaction between two entities (e_i and e_j) is defined by the combination of both their functions:

$$e_i \leftrightarrow e_j \quad := \quad f_{e_i+e_j} = (f_{e_i} + f_{e_j}). \tag{7.4}$$

Since an observer o is also an entity, we can define it as

$$o_i \quad := \quad e_j \quad | \quad f_{e_i+e_j}(s_{e_i}) = s'_{e_i} \neq s_{e_i}, \text{ where } i \text{ can equal } j, \tag{7.5}$$

which reads "an observer is an entity that, given an interaction with another entity, changes its state." The existence of an observer becomes a question of the property of

an entity's state: an entity's state must be posed to be altered by the state of another entity in some way via an interaction. I was explicit in making sure an observer is allowed to interact with itself, which intuitively includes things like the observer as a whole being an internal process, although this is certainly up for debate. In addition, an entity $e = e_1 \cup e_2 \cup ...e_n$ can interact with the entities it is composed of, such as e_1. This framework allows for any entity, made of entities, to interact with any other entity, made of entities.

For fun, let's explore the limits of such entities:

$$\lim_{|e|\to\infty} \{e\} = \text{Universe}, \tag{7.6}$$

that is to say, the largest possible entity is the Universe. Furthermore, the state of the Universe dictates how it changes. In the other direction:

$$\lim_{|e|\to 0} \{e\} = 0, \tag{7.7}$$

meaning an entity composed of no entities is non-existent, an empty set.

Although this exercise has been in no means rigorous, it serves as a mental road map for thinking about the world in a way that allows the emergence of living things. Notions such as observers, levels of organization, boundaries, individuals, etc could be expressed using a self-referential framework. It could be worth pursuing the development of such a model to see if emergent functions like experience, meaning, and cooperation are possible. However, it still remains unclear that these properties of everyday life could be captured in such a model because of its tractability. How is it possible to model an entity on a mammalian level with this framework? Given our current implementation of programs and data in the context of computers, this may not be possible to explore using modern technology. Because modern computers are rooted in the idea of applying programs to sets of data rather than fusing programs

and data together within the same physical state, living systems may quite simply impossible to understand under our current computational framework.

7.1 Towards a New Model of the World

We currently have few insights into how living systems might *quantifiably* differ from their non-living counterparts, as in a mathematical foundation to explain away our observations of evolution, emergence, innovation, and organization. Development of a theory of living systems, if at all possible, demands mathematical understanding of how data is generated, collected, and changes over time, not unlike current well-established scientific disciplines. After all, living systems are comprised of physical elements: atoms, charges, and masses. It seems that it is only a matter of time before someone comes along and fits a mathematical model that bridges the gap between our observations of living systems and current scientific models. Given a collection of observations on some level of scale–chemical, quantum, astronomical–a mathematical model is invented to explain and, more importantly, predict those observations. Through this process, most scientific disciplines are formed.

According to many philosophers who agree with Alfred N. Whitehead, the folly of these various (largely) disconnected scientific disciplines is the omission of perfectly valid observations and measurements in order to satisfy a rigid mathematical approach. Human thoughts such as "What do yellow tulips mean?" and "What was I thinking when I ordered that nasty taco" have no place in current mathematical models. Although Fourier transformations can be easily applied to classical music in order to analyze which frequencies are most common, it lends no understanding to how it makes a person *feel* or what thoughts it elicits to the listener. The elimination of such

details (details that are common and entirely vital in living systems) is hindering our understanding of what life is and what not-life is. As Whitehead is so insistent about, the mere observation of facts as positivists is a nonsensical approach to understanding Nature. Why should scientists only be concerned with plain descriptions of Nature without probing their underlying meanings? Common criticisms to this claim include using teleology, or giving 'will' or meaning to objects. Here, it is important to draw a boundary where 'meaning' is being applied: Studying meaning does not imply that all objects have a 'meaning', such as the 'meaning of an electron moving through a magnetic field.' Instead, the idea of 'meaning' is given to living entities, much like we humans draw meaning from printed symbols. An RNA molecule has a well-defined function in the appropriate context of a cell, which is not to exclude 'meaning of the RNA to the cell'. **Subjectivity, rather than pure objectivity, of the laws of nature should be emphasized.**

Current mathematical models that describe the precession of physical entities thorough space-time, or any combination of such entities, are not sufficient to understand the phenomenon that are more common to experiencing everyday life as humans (Whitehead 1927, Whitehead 1928, and Whitehead 1934). And yet, we remain largely ignorant to the importance of experience and sensing. I want to stress that although these terms are generally used to describe the human experience, it should not exclude the most general notions of experience and sense as possible. For example, how does a cell 'sense' RNA? Because of this failure, and perhaps others, we currently lack a coherent theory general enough to encompass important aspects of life, which are the crux of commonplace thoughts and feelings among humans that drive our societal, and possible other biological, processes.

REFERENCES

Abbot, E Stanley. 1916. "The Causal Relations between Structure and Function in Biology." *The American Journal of Psychology* 27 (2): 245–250.

Abbott, Derek. 2013. "The Reasonable Ineffectiveness of Mathematics." *Proceedings of the IEEE* 101 (October): 2147–2153. doi:10.1109/JPROC.2013.2274907.

Adams, Alyssa M, Angelica Berner, Paul C W Davies, and Sara I Walker. 2017. "Physical Universality, State-Dependent Dynamical Laws and Open-Ended Novelty." *Entropy* 19 (9).

Adams, Alyssa M, and Sara Imari Walker. 2017. "Real-world Open-ended Evolution: A League Of Legends Adventure." *International Journal of Design & Nature and Ecodynamics* 12 (4): 11.

Adams, Alyssa, Hector Zenil, Paul C. W. Davies, and Sara Imari Walker. 2017. "Formal Definitions of Unbounded Evolution and Innovation Reveal Universal Mechanisms for Open-Ended Evolution in Dynamical Systems." *Scientific Reports* 7 (1): 997. doi:10.1038/s41598-017-00810-8.

Ahn, Yong-Yeol, James P. Bagrow, and Sune Lehmann. 2010. "Link communities reveal multiscale complexity in networks." *Nature* 466, no. 7307 (August): 761–764. http://dx.doi.org/10.1038/nature09182.

Allen, J. A., and B. C. and Clarke. 1984. "Frequency-dependent selection- homage to Poulton." *Biological Journal of the Linnean Society* 23:15–18.

Austin, D. 2017. "How Google Finds Your Needle in the Web's Haystack." http://www.ams.org/samplings/feature-column/fcarc-pagerank.

Babbage, Benjamin Herschel. 1864. "Portion of Babbage's difference engine." *Harper's new monthly magazine* 30 (175).

Banzhaf, Wolfgang, Bert Baumgaertner, and Guillaume et al Beslon. 2016. "Defining and simulating open-ended novelty: requirements, guidelines, and challenges." *Theory in Biosciences* 135 (3): 1–31. doi:10.1007/s12064-016-0229-7.

Bedau, M A, and N H Packard. 1992. "Measurement of Evolutionary Activity, Teleology, and Life." In *Artificial Life II, Santa Fe Institute Studies in the Sciences of Complexity,* edited by In C Langton, C Taylor, D Farmer, and S Rasmussen. M. A. Bedau / N. H. Packard.

Bedau, Mark. 1991. "Can biological teleology be naturalized?" *The Journal of Philosophy* 88, no. 11 (November): 647–655.

Bedau, Mark A, John S McCaskill, Norman H Packard, Steen Rasmussen, Chris Adami, David G Green, Takashi Ikegami, Kunihiko Kaneko, and Thomas S Ray. 2000. "Open problems in artificial life." *Artificial Life* 6 (4): 363–376.

Bedau, Mark A, Emile Snyder, C Titus Brown, Norman H Packard, et al. 1997. "A comparison of evolutionary activity in artificial evolving systems and in the biosphere." In *Fourth European Conference on Artificial life,* 125–134. Citeseer, MIT Press, Cambridge, MA.

Bettencourt, Luis MA, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B West. 2007. "Growth, innovation, scaling, and the pace of life in cities." *Proceedings of the national academy of sciences* 104 (17): 7301–7306.

Bianco, Raffaele, and Stefano Nolfi. 2004. "Toward open-ended evolutionary robotics: evolving elementary robotic units able to self-assemble and self-reproduce." *Connection Science* 16 (4): 227–248.

"Big Data." 2015. ibm.com/software/data/bigdata/.

Bornholdt, Stefan, and Thimo Rohlf. 2000. "Topological Evolution of Dynamical Networks: Global Criticality from Local Dynamics." *Physical Review Letters* 84, no. 26 (June): 6114–6117. https://link.aps.org/doi/10.1103/PhysRevLett.84.6114.

Borriello, Enrico, and Sara Imari Walker. 2017. "An Information Theoretic Classification of Elementary Cellular Automata." *Complexity.*

Box, G. E. P. 1976. "Science and Statistics." *Journal of the American Statistical Association* 71.

Boyd, A.B., and J.P. Crutchfield. 2016. "Maxwell demon dynamics: Deterministic chaos, the Szilard map, and the intelligence of thermodynamic systems." *Physical review letters* 116 (19).

Brender, Noah Moss. 2012. "Sense-Making and Symmetry-Breaking: Merleau-Ponty, Cognitive Science, and Dynamic Systems Theory." 2012 CSCP Graduate Essay Prize Winner.

Breukelaar, R., and T Bäck. 2005. "Using a genetic algorithm to evolve behavior in multi dimensional cellular automata: emergence of behavior," 107–114. Washing-

ton, DC, USA: 7th Annual Conference on Genetic and Evolutionary Computation (GECCO '05).

Briggs, James. 2016. "Babbage's Difference Engine at Computer History Museum." http://www.jebriggs.com/blog/2016/01/babbages-analytical-engine-at-computer-history-museum/.

Buchanan, Andrew, Norman H Packard, and Mark A Bedau. 2011. "Measuring the evolution of the drivers of technological innovation in the patent record." *Artificial Life* 17 (2): 109–122.

Buiatti, Marcello, and G. Longo. 2013. "Randomness and Multi-level Interactions in Biology." *Theory of Biosciences* 132 (3): 139–158.

Calude, C. S., and G. Longo. 2016. "The Deluge of Spurious Correlations in Big Data." *Foundations of Science:* 1–18.

captainhowdy27. 2016. "Clever latte art." https://imgur.com/gallery/uFoUjeQ.

"Cassiopeia." 2017. github.com/meraki-analytics/cassiopeia.

Chalmers, David. 1996. "Does a Rock Implement Every Finite-State Automaton?" *Synthese* 108.

Chalmers, Devin W, C Cooper Francis, Noah M Pepper, and Mark A Bedau. 2010. "High-content Words in Patent Records Reflect Key Innovations in the Evolution of Technology." In *Alife XII Conference,* 838–845. Artificial Life 12, MIT Press, Cambridge, MA.

Cong, Le, F Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D Hsu, et al. 2013. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science (New York, N.Y.)* 339, no. 6121 (February): 819–823. doi:10.1126/science.1231143.

Conway, J. 1970. "The Game of Life." *Scientific American.*

Cook, Matthew. 2004. "Universality in elementary cellular automata." *Complex Systems* 15 (1): 1–40.

Corominas-Murtra, Bernat, Jordi Fortuny, and Ricard V. Solé. 2014. "Towards a mathematical theory of meaningful communication." *Scientific Reports* 4. http://dx.doi.org/10.1038/srep04587.

Corominas-Murtra, Bernat, Luis Seoane, and Ricard Solé. 2016. "Zipf's law, unbounded complexity and open-ended evolution." Https://arxiv.org/abs/1612.01605.

Crutchfield, J.P. 1994. "The calculi of emergence: computation, dynamics and induction." *Physica D: Nonlinear Phenomena* 75 (1).

Danchin, Antoine. 2009. "Bacteria as computers making computers." *FEMS Microbiology Reviews* 33 (1): 3–26.

Das, R., M. Mitchell, and J. P. Crutchfield. 1994. "A genetic algorithm discovers particle computation in cellular automata." In *Lecture Notes in Computer Science,* 866:344–353. Third Conference on Parallel Problem Solving from Nature, Springer.

"Datreant: Persistent, pythonic trees for heterogeneous data." 2017. http://datreant.readthedocs.io/en/latest/.

Davies, Paul C W, and Sara Imari Walker. 2016. "The Hidden Simplicity of Biology: A Key Issues Review." *Reports on Progress in Physics* 79 (10): 102601.

DeDeo, Simon. 2011. "Collective Phenomena and Non-Finite State Computation in a Human Social System." *PLoS ONE* 9 (6): 109–122.

Deutsch, David. 2013. "Constructor theory." *Synthese* 190 (18): 4331–4359. doi:10.1007/s11229-013-0279-z.

———. 2016. "The logic of experimental tests, particularly of Everettian quantum theory." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 55:24–33.

Deutsch, David, and Chiara Marletto. 2014. "Constructor theory of information." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* 471, no. 2174 (December). http://rspa.royalsocietypublishing.org/content/471/2174/20140540.abstract.

Draves, Scott. 2017. "Electric Sheep: CROWDSOURCED EVOLVING ART." https://gold.electricsheep.org/license/.

Du, K.L., and M.N.S. Swamy. 2013. *Neural Networks and Statistical Learning.* Springer London. https://books.google.com/books?id=Veo%5C_nwEACAAJ.

Ellis, George FR. 2011. "Top-down causation and emergence: some comments on mechanisms." *Interface Focus.*

Fernando, Chrisantha, George Kampis, and Eörs Szathmáry. 2011. "Evolvability of natural and artificial systems." In *2nd European Future Technologies Conference and Exhibition 2011,* 7:73–76. FET 11, Elsevier.

Fraczkowski, Rachel, and Megan Olsen. 2014. "An Agent-Based Predator-Prey Model with Reinforcement Learning." In *Swarmfest 2014: 18th Annual Meeting on Agent-Based Modeling & Simulation.* June.

Frank, S. A. 2009. "Natural selection maximizes Fisher information." *Journal of Evolutionary Biology* 22 (2): 231–244. doi:10.1111/j.1420-9101.2008.01647.x.

Friston, K. 2013. "Life As We Know It." *Royal Society Publishing* 10 (86).

Games, Riot. 2016a. "Player Numbers." riotgames.com/tags/player-numbers.

———. 2016b. "Stats Straight from the Source." developer.riotgames.com/.

———. 2017. "Getting Started." https://developer.riotgames.com/getting-started.html.

Gardner, Martin. 1970. "Mathematical Games – The fantastic combinations of John Conway's new solitaire game "life"." *Scientific American* 223.

GLAVIN, Daniel P., Andrew D. AUBREY, Michael P. CALLAHAN, Jason P. DWORKIN, Jamie E. ELSILA, Eric T. PARKER, Jeffrey L. BADA, Peter JENNISKENS, and Muawia H. SHADDAD. 2010. "Extraterrestrial amino acids in the Almahata Sitta meteorite." *Meteoritics & Planetary Science* 45 (10-11): 1695–1709. doi:10.1111/j.1945-5100.2010.01094.x.

Glavin, Daniel P., Jeffrey L. Bada, Karen L. F. Brinton, and Gene D. McDonald. 1999. "Amino acids in the Martian meteorite Nakhla." *Proceedings of the National Academy of Sciences* 96, no. 16 (August): 8835–8838.

Gödel, Kurt. 1931. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I." *Monatshefte für Mathematik und Physik* 38 (1): 173–198.

Goel, N. S. 1971. "On the Volterra and Other Non-Linear Models of Interacting Populations." *Academic Press.*

Goldenfeld, Nigel, and Carl Woese. 2007. "Biology's next revolution." *Nature* 445, no. 7126 (January): 369–369. http://dx.doi.org/10.1038/445369a.

———. 2011. "Life is Physics: Evolution as a Collective Phenomenon Far From Equilibrium." *Annual Review of Condensed Matter Physics* 2, no. 1 (November): 375–399.

Gordon, Whitson. 2012. "Silence Noisy Neighbors by Transmitting Your Music to Their Speakers." October. https://lifehacker.com/5852903/silence-noisy-neighbors-by-transmitting-signals-through-their-own-speakers.

Grandjean, Martin. 2017. "Digital Humanities on Twitter." https://commons.wikimedia.org/wiki/File:Digital_Humanities_on_Twitter.png.

Griffiths, Paul E. 2017. "Genetic, epigenetic and exogenetic information in development and evolution." *Interface Focus* 7 (5). doi:10.1098/rsfs.2016.0152. eprint: http://rsfs.royalsocietypublishing.org/content/7/5/20160152.full.pdf.

Griffiths, Paul E., Arnaud Pocheville, Brett Calcott, Karola Stotz, Hyunju Kim, and Rob Knight. 2015. "Measuring Causal Specificity," 82 (4): 529–555. doi:10.1086/682914.

Group, EvCA. 2000. "Evolving Cellular Automata with Genetic Algorithms." http://csc.ucdavis.edu/~evca/Projects/evca.html.

Guttenberg, Nicholas, and Nigel Goldenfeld. 2008. "Cascade of complexity in evolving predator-prey dynamics." *Physical Review Letters* 100, no. 5 (February): 058102.

Haarlem, Cornelis Cornelisz. 1604. *Antrum Platonicum.* Print.

Haldane, J. B. S. 1947. *What is Life?* New York, Boni / Gaer.

Hansen, Drew. 2016. "How To Become A Data Scientist In 2017." https://www.forbes.com/sites/drewhansen/2016/10/21/become-data-scientist/#947bbe187d33.

Hayes-Roth, Frederick, Donald Waterman, and Douglas Lenat. 1983. *Building Expert Systems.* Addison-Wesley.

Hickinbotham, S. J., and S. Stepney. 2016. "Bio-Reflective Architectures for Evolutionary Innovation." In *Proceedings of the Artificial Life Conference,* edited by MIT Press.

Hoel, Eric P. 2017. "When the Map Is Better Than the Territory." *Entropy* 19 (5).

Hofstadter, D. R. 1979. *Gödel, Escher, Bach: an Eternal Golden Braid.* Basic Books, Inc.

Hooft, G. 2014. "The Cellular Automaton Interpretation of Quantum Mechanics." *arXiv:1405.1548.*

Hordijk, W. 2013. "The EvCA project: a brief history." *Complexity* 18 (5): 15–19.

Hoyle, Fred. 1957. *The Black Cloud.* William Heinemann Ltd.

Hsu, Jeremy. 2015. "Inside the Largest Virtual Psychology Lab in the World." https://backchannel.com/inside-the-largest-virtual-psychology-lab-in-the-world-7c0d2c43cda5.

Inc., Pandora Media. 2017. "About The Music Genome Project®." https://www.pandora.com/about/mgp.

Israeli, N., and N. Goldenfeld. 2006. "Coarse-graining of cellular automata, emergence, and the predictability of complex systems." *Phys. Rev. E* 73.

Janzing, D. 2010. "Is There a Physically Universal Cellular Automaton or Hamiltonian?" *ArXiv:1009.1720.*

Jaynes, E T. 2003. *Probability Theory: The Logic of Science.* Edited by L G Bretthorst. Cambridge University Press.

Jeong, et al., H. 2000. "The large-scale organization of metabolic networks." *Nature* 407.

Karpathy, Andrej. 2017. "CS231n Convolutional Neural Networks for Visual Recognition." http://cs231n.github.io/neural-networks-1/.

Kataoka, N., and K. Kaneko. 2000. "Natural language from function dynamics." *Biosystems* 10 (1-2): 1–11.

Kataoka, N., and K. Kunihiko. 2001. "Functional Dynamics II: Syntactic Structure." *Physica D: Nonlinear Phenomena.*

Kataoka, Naoto, and Kunihiko Kaneko. 2000. "Functional dynamics I: Articulation process." *Physica D: Nonlinear Phenomena* 138, nos. 3-4 (April): 225–250.

Kauffman, S.A. 2000. *Investigations.* Oxford University Press.

Kauffman, Stuart A. 2011. "There Are More Uses For A Screwdriver Than You Can Calculate." http://www.wbur.org/npr/135113346/there-are-more-uses-for-a-screwdriver-than-you-can-calculate.

Kauffman, Stuart A. 2016. *Humanity in a Creative Universe.* Oxford University Press.

Kim, Hyunju, Paul Davies, and Sara Imari Walker. 2015. "New scaling relation for information transfer in biological networks." *Journal of The Royal Society Interface* 12, no. 113 (December). http://rsif.royalsocietypublishing.org/content/12/113/20150944.abstract.

Knott, Cargill Gilsto. 1911. *Quote from undated letter from Maxwell to Tait.* Letter.

Koga, Toshiki, and Hiroshi Naraoka. 2017. "A new family of extraterrestrial amino acids in the Murchison meteorite." *Scientific Reports* 7 (1): 636. doi:10.1038/s41598-017-00693-9.

Kontoleon, Nectarios, Lucia Falzon, and Philippa Pattison. 2013. "Algebraic structures for dynamic networks." *Journal of Mathematical Psychology* 57 (6): 310–319.

Kühnlenz, Florian, and Pedro H. J. Nardelli. 2016. "Dynamics of Complex Systems Built as Coupled Physical, Communication and Decision Layers." *PLOS ONE* 11, no. 1 (January): 1–19. doi:10.1371/journal.pone.0145135.

Landauer, Rolf. 1961. "Irreversibility and heat generation in the computing process." *IBM Journal of Research and Development* 5 (3): 183–191.

Langton, C. G. 1984. "Self-reproduction in cellular automata." *Physics D: Nonlinear Phenomena* 10 (1-2): 135–144.

Langton, C.G. 1990. "Computation at the edge of chaos: phase transitions and emergent computation." *Physica D: Nonlinear Phenomena* 42.

Laubichler, Manfred D, and Jürgen Renn. 2015. "Extended evolution: A conceptual framework for integrating regulatory networks and niche construction." *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 324, no. 7 (November): 565–577.

Levary, David, Jean-Pierre Eckmann, Elisha Moses, and Tsvi Tlusty. 2012. "Loops and self-reference in the construction of dictionaries." *Physical Review X* 2, no. 3 (September): 031018.

Levin, M. 2012. "Morphogenetic fields in embryogenesis, regeneration and cancer: non-local control of complex patterning." *Biosystems* 109.

Life, Conway's Game of. 2017. "Lipa, Chris." http://www.math.cornell.edu/~lipa/mec/lesson6.html.

Ling, Hong, Sandhya Samarasinghe, and Don Kulasiri. 2013. "Novel recurrent neural network for modelling biological networks: Oscillatory p53 interaction dynamics." *Biosystems* 114:191–205.

Lobo, D, W S Beane, and M Levin. 2012. "Modeling Planarian Regeneration: A Primer for Reverse-Engineering the Worm." *PLoS Computation Biology.*

Lobo, D, and M Levin. 2015. "Inferring Regulatory Networks from Experimental Morphological Phenotypes: A Computational Method Reverse-Engineers Planarian Regeneration." *PLoS Computational Biology* 11 (6).

Longo, G., P A Miquel, C Sonnenschein, and A M Soto. 2012. "Is information a proper observable for biological organization?" *Progress in Biophysics and Molecular Biology* 109:108–114.

Loreto, Vittorio, Vito D. P. Servedio, and Steven H. Strogatz. 2016. "Dynamics on Expanding Spaces: Modeling the Emergence of Novelties." *Creativity and Universality in Language:* 59–83.

Lotka, A. J. 1910. "Contribution to the Theory of Periodic Reaction." *Journal of Physical Chemistry* 14 (3).

Lovinger, Justin, Iren Valova, MacKenzie Rogers, Ryan Nadeau, and Natacha Gueorguiev. 2014. "Harnessing Mother Nature: Optimizing Genetic Algorithms for Adaptive Systems." *Procedia Computer Science* 36:523–528.

Lui, Ting L, Terrazas G, Hector Zenil, Cameron Alexander, and Natalio Krasnogor. 2015. "Complexity Measurement Based on Information Theory and Kolmogorov Complexity." In *Artificial Life,* 205–24. MIT Press, Cambridge, MA.

Ma, Shouwen, Andries Ter Maat, and Manfred Gahr. 2017. "Power-law scaling of calling dynamics in zebra finches." *Scientific Reports* 7 (1): 8397. doi:10.1038/s41598-017-08389-w.

Maccone, Lorenzo. 2009. "Quantum Solution to the Arrow-of-Time Dilemma." *Physical Review Letters* 103.

Magliocco, Mario, Steven Glaser, and Timothy Kneafsey. 2013. "Laboratory Investigation of Supercritical $CO_2 Use in Geothermal Systems$." In *PROCEEDINGS, Thirty-Eighth Workshop on Geothermal Reservoir Engineering.*

Mandal, D., and C. Jarzynski. 2012. "Work and information processing in a solvable model of Maxwell's demon." *Proceedings of the National Academy of Sciences* 109 (29).

Margolus, N. 1984. "Physics-like Models of Computation." *Physica D: Nonlinear Phenomena.*

Marletto, Chiara. 2015. "Constructor theory of life." *Journal of The Royal Society Interface* 12, no. 104 (January). http://rsif.royalsocietypublishing.org/content/12/104/20141226.abstract.

———. 2016. "Constructor Theory of Thermodynamics." *(Preprint) arXiv:1608.02625.*

Marzen, Sarah E., and Simon DeDeo. 2017. "The evolution of lossy compression." *Journal of The Royal Society Interface* 14, no. 130 (May). http://rsif.royalsocietypublishing.org/content/14/130/20170166.abstract.

Medernach, David, Jeannie Fitzgerald, Simon Carrignon, and Ryan Conor. 2015. "Evolutionary Progress in Heterogenous Cellular Automata (HetCA)." In *European Conference on Artificial Life 2015,* 512–519. Citeseer, July.

Mitchell, Melanie. 2011. "What is Computation?" *Ubiquity, an ACM Publication.*

Mitchell, Melanie, James P Crutchfield, and Peter T Hraber. 1994. "Evolving cellular automata to perform computations: Mechanisms and impediments." *Physica D: Nonlinear Phenomena* 75, nos. 1-3 (August): 361–391.

Montgomery, R. 1993. "Gauge Theory of the Falling Cat." *Dynamics and Control of Mechanical Systems:* 193–218.

Moore, Douglas, Sara I Walker, and Michael Levin. 2017. "Cancer as a disorder of patterning information: computational and biophysical perspectives on the cancer problem." *Convergent Science Physical Oncology* 3 (4): 043001. http://stacks.iop.org/2057-1739/3/i=4/a=043001.

Mora, Thierry, Aleksandra M. Walczak, Lorenzo Del Castello, Francesco Ginelli, Stefania Melillo, Leonardo Parisi, Massimiliano Viale, Andrea Cavagna, and Irene Giardina. 2016. "Local equilibrium in bird flocks." *Nature Phys* 12, no. 12 (December): 1153–1157. http://dx.doi.org/10.1038/nphys3846.

Müller, Johann Helfrich. 1786. *Description of his newly invented computing machine.* Varrentrapp son / Wenner.

Museum, Computer History. 2017. "The Babbage Engine." http://www.computerhistory.org/babbage/.

Myhill, J. 1963. "The Converse of Moore's Garden-of-Eden Theorem." In *Proceedings of the American Mathematical Society 14.*

Nat. 2010. "BOIDS + PERLIN NOISE FLOW FIELD." https://flashyprogramming.wordpress.com/2010/03/22/boids-perlin-noise-flow-field/.

Nemenman, Ilya. 2012. *Information theory and adaptation.* Edited by Michael E. Wall. CRC Press.

Newman, MEJ. 2005. "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics* 46, no. 5 (September): 323–351. doi:10.1080/00107510500052444.

Nghe, Philippe, Wim Hordijk, Stuart A. Kauffman, Sara I. Walker, Francis J. Schmidt, Harry Kemble, Jessica A. M. Yeates, and Niles Lehman. 2015. "Prebiotic network evolution: six key parameters." *Molecular BioSystems* 11 (12): 3206–3217. doi:10.1039/C5MB00593K.

Nobili, R., and U. Pesavento. 1994. "John von Neumann's automata revisited." *Artificial Worlds and Urban Studies.*

Noble, Denis. 2012. "A theory of biological relativity: no privileged level of causation." *Interface Focus* 2 (1): 55–64.

Nurse, Paul. 2008. "Life, logic and information." *Nature Horizons* 454 (24): 424–426.

Oizumi, Masafumi, Larissa Albantakis, and Giulio Tononi. 2014. "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0." *PLoS Computation Biology.*

Oka, Mizuki, Yasuhiro Hashimoto, and Takashi Ikegami. 2015. "Open-ended evolution in a web system." In *Late Breaking Papers at the European Conference on Artificial Life.*

Oka, Mizuki, and Takashi Ikegami. 2013. "Exploring Default Mode and Information Flow on the Web." *PLoS ONE* 8 (4).

Pavlic, Theodore P, Alyssa M Adams, Paul C W Davies, and Sara Imari Walker. 2014. "Self-referencing cellular automata: A model of the evolution of information control in biological systems." In *Proc. of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems,* edited by J. Pollack, M.

Bedau, P. Husbands, T. Ikegami, and R. Watson. Artificial Life 14, MIT Press, Cambridge, MA.

"PiAndWhippedCream", User. 2008. "Rule 30 2000 Generations." https://commons.wikimedia.org/wiki/File:Rule_30_2000Generations.png.

Plato. 1943. *Plato's The Republic.* New York Books, Inc.

Rendell, Paul. 2000. "This is a Turing Machine implemented in Conway's Game of Life." http://rendell-attic.org/gol/tm.htm.

Rex, Andrew. 2017. "Maxwell's Demon—A Historical Review." *Entropy* 19 (6).

Reynolds, Craig. 1987. "Flocks, herds and schools: A distributed behavioral model." In *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques,* 25–34. Association for Computing Machinery.

"Riot Manifesto." 2017. riotgames.com/riot-manifesto.

Rosen, Robert J. 1981. "Pattern Generation in Networks." Academic Press.

Ruiz-Mirazo, Kepa, and Alvaro Moreno. 2012. "Autonomy in evolution: from minimal to complex life." *Synthese* 185 (1): 21–52. doi:10.1007/s11229-011-9874-z.

Ruiz-Mirazo, Kepa, Juli Peretó, and Alvaro Moreno. 2004. "A universal definition of life: autonomy and open-ended evolution." *Origins of Life and Evolution of the Biosphere* 34 (3): 323–346. doi:10.1023/B:ORIG.0000016440.53346.dc.

Ruiz-Mirazo, Kepa, Jon Umerez, and Alvaro Moreno. 2008. "Enabling conditions for 'open-ended evolution'." *Biology & Philosophy* 23 (1): 67–85. doi:10.1007/s10539-007-9076-8.

Salo, V., and I. Törmä. 2017. "A One-Dimensional Physically Universal Cellular Automaton." In *Conference on Computability in Europe.* Springer, Cham.

Schaeffer, L. 2015. "A Physically Universal Celllular Automaton." In *Proceedings of the 2015 Conferences on Innovations in Theoretical Computer Science.* ACM.

Schiff, Joel L. 2011. *Cellular Automata: A Discrete View of the World.* Wiley & Sons Inc.

Schlitt, T., and A. Brazma. 2007. "Current approaches to gene regulatory network modelling." *BMC Bioinformatics* 8 (6).

Schrödinger, E. "What is life?" 1944. Cambridge University Press.

Schultz, Jeff. 2016. "How Much Data is Created on the Internet Each Day?" September. https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/.

Sekara, Vedran, Arkadiusz Stopczynski, and Sune Lehmann. 2016. "Fundamental structures of dynamic social networks." *Proceedings of the National Academy of Sciences* 113, no. 36 (September): 9977–9982.

Seoane, Luis F, and Ricard Solé. 2017. "Information theory, predictability, and the emergence of complex life." *(Preprint) arXiv:1701.02389.*

"SergioJimenez", User. 2015. "Zipf 30wiki en labels." https://commons.wikimedia.org/wiki/File:Zipf_30wiki_en_labels.png.

Seyfarth, Robert M, Dorothy L Cheney, and Thore J Bergman. 2005. "Primate social cognition and the origins of language." *Trends in cognitive sciences* 9 (6): 264–266.

Shannon, Claude E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (3).

Shwartz-Ziv, Ravid, and Naftali Tishby. 2017. "Opening the Black Box of Deep Neural Networks via Information." *(Preprint) arXiv:1703.00810.*

Skusa, Andre, and Mark A Bedau. 2002. "Towards a comparison of evolutionary creativity in biological and cultural evolution." In *Eighth international conference on Artificial life,* 8:233–242. ICAL 2003, MIT Press, Cambridge, MA.

Smilga, Nir. 2014. "Starling Dance." http://www.pikiwiki.org.il/?action=gallery&img_id=38516.

Smith, Eric. 2008. "Thermodynamics of natural selection I: Energy flow and the limits on organization." *Journal of Theoretical Biology* 252 (2): 185–197.

Soros, LB. 2016. "This page contains visual results from the ALife XV paper How the Strictness of the Minimal Criterion Impacts Open-Ended Evolution." http://eplex.cs.ucf.edu/chromaria-resources/alife16/.

Soros, LB, and Kenneth O Stanley. 2014. "Identifying necessary conditions for open-ended evolution through the artificial life world of Chromaria." In *Artificial Life Conference (ALife 14),* 793–800. Artificial Life 14, MIT Press, Cambridge, MA.

Stopczynski, Arkadiusz, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. 2014. "Measuring Large-Scale Social Networks with High Resolution." *PLoS ONE* 9 (4).

Szalay, K. Z., and P. Csermely. 2013. "Perturbation Centrality and Turbine: A Novel Centrality Measure Obtained Using a Versatile Network Dynamics Tool." *PLoS ONE* 8 (10).

Tassi, Paul. 2016. "Riot Games Reveals 'League of Legends' Has 100 Million Monthly Players." September. https://www.forbes.com/sites/insertcoin/2016/09/13/riot-games-reveals-league-of-legends-has-100-million-monthly-players/#353676ba5aa8.

Taylor, Tim. 2004. "Redrawing the boundary between organism and environment." In *Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems,* edited by J. Pollack, M. Bedau, P. Husbands, T. Ikegami, and R. Watson, 268–273. Artificial Life 9.

Taylor, Tim, Mark Bedau, Alastair Channon, David Ackley, Wolfgang Banzhaf, Guillaume Beslon, Emily Dolson, Tom Froese, Simon Hickinbotham, Takashi Ikegami, et al. 2016. "Open-Ended Evolution: Perspectives from the OEE1 Workshop in York." *Artificial Life* 22, no. 3 (August): 408–423.

Taylor, Timothy J. 1999. "From artificial evolution to artificial life." PhD diss., University of Edinburgh. College of Science and Engineering. School of Informatics.

Tejedor, Alejandro, Anthony Longjas, Ilya Zaliapin, Samuel Ambroj, and Efi Foufoula-Georgiou. 2017. "Network robustness assessed within a dual connectivity framework: joint dynamics of the Active and Idle Networks." *Scientific Reports* 7 (1): 8567. doi:10.1038/s41598-017-08714-3.

The Beginning of Infinitiy: Explanations That Transform the World. 2011. Penguin UK.

"The Online Algorithmic Complexity Calculator." 2017. http://complexitycalculator.com/HowItWorks.html.

Tishby, Naftali, Fernando C. Pereira, and William Bialek. 1999. "The Information Bottleneck Method." In *The 37th annual Allerton Conference on Communication, Control, and Computing,* 368–377.

Tisseur, Pierre. 2000. "Cellular automata and Lyapunov exponents." *Nonlinearity* 13, no. 5 (March): 1547.

Toffoli, T. 1984. "Cellular automata as an alternative to (rather than an approximation of) differential equations in modeling physics." *Physica D: Nonlinear Phenomena* 10.

Tononi, Giulio, Melanie Boly, Marcello Massimini, and Christof Koch. 2016. "Integrated information theory: from consciousness to its physical substrate." *Nat Rev Neurosci* 17, no. 7 (July): 450–461. http://dx.doi.org/10.1038/nrn.2016.44.

Transtrum, Mark K., Benjamin B. Machta, Kevin S. Brown, Bryan C. Daniels, Christopher R. Myers, and James P. Sethna. 2015. "Perspective: Sloppiness and emergent theories in physics, biology, and beyond," *The Journal of Chemical Physics* 143 (1): 010901. doi:10.1063/1.4923066.

"TRAPPIST-1." 2017. http://www.trappist.one/.

Turing, Alan. 1937. "On Computable Numbers With an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 42 (1).

Vicsek, Tamás, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. 1995. "Novel Type of Phase Transition in a System of Self-Driven Particles." *Physical Review Letters* 75, no. 6 (August): 1226–1229. https://link.aps.org/doi/10.1103/PhysRevLett.75.1226.

Vigen, Tyler. 2017. "Spurious Correlations." http://www.tylervigen.com/spurious-correlations.

Villarreal, Luis. 2008. "Are Viruses Alive?" https://www.scientificamerican.com/article/are-viruses-alive-2004/.

Vladar, Harold P. de, and Nicholas H. Barton. 2011. "The contribution of statistical physics to evolutionary biology." *Trends in Ecology & Evolution* 26 (8): 424–432. doi:https://doi.org/10.1016/j.tree.2011.04.002.

Vladar, Harold P. de, Mauro Santos, and Eörs Szathmáry. 2017. "Grand Views of Evolution," *Trends in Ecology & Evolution* 32, no. 5 (September): 324–334. doi:10.1016/j.tree.2017.01.008.

Von Neumann, J. 1966. "Theory of Self-Reproducing Automata." *University of Illinois Pres.*

Wagner, Andreas. 2014. *Arrival of the Fittest: Solving Evolution's Greatest Puzzle.* Penguin.

Walker, Sara I. 2015. "Is Life Fundamental?" In *Questioning the Foundations of Physics: Which of our Fundamental Assumptions are Wrong?,* edited by A Aguirre, B Foster, and Z Merali. Springer.

Walker, Sara Imari. 2014. "Top-down causation and the rise of information in the emergence of life." *Information* 5 (3): 424–439.

———. 2015. "The Descent of Math." In *The Mysterious Connection Between Physics and Mathematics?,* edited by A Aguirre, B Foster, and Z Merali. Springer.

Walker, Sara Imari, Benjamin J. Callahan, Gaurav Arya, J. David Barry, Tanmoy Bhattacharya, Sergei Grigoryev, Matteo Pellegrini, Karsten Rippe, and Susan M. Rosenberg. 2013. "Evolutionary dynamics and information hierarchies in biological systems." *Annals of the New York Academy of Sciences* 1305 (1): 1–17. doi:10.1111/nyas.12140.

Walker, Sara Imari, Luis Cisneros, and Paul CW Davies. 2012. "Evolutionary transitions and top-down causation." In *Artificial Life 13,* edited by Chris Adami, David Bryson, Charles Ofria, and Robert Pennock, 283–290. Artificial Life, MIT Press, Cambridge, MA.

Walker, Sara Imari, and Paul C W Davies. 2013. "The algorithmic origins of life." *Journal of the Royal Society Interface* 10 (79): 20120869.

Walker, Sara Imari, Paul C. W. Davies, and George F. R. Ellis. 2017. *From Matter to Life: Information and Causality.* Cambridge: Cambridge University Press. doi:DOI:10.1017/9781316584200.

Watts, D. J., and S. H. Strogatz. 1998. "Collective dynamics of 'small-world' networks." *Nature* 393.

Weisstein, Eric W. 2017. "Rule 110." http://mathworld.wolfram.com/Rule110.html.

Wheeler, John A. 1990. *Information, Physics, Quantum: The Search for Links.* Physics Dept., University of Texas.

Whitehead, Alfred North. 1927. *Symbolism, its Meaning and Effect.* Barbour-Page Lectures, April.

———. 1928. *Process and Reality, an Essay in Cosmology.* Gifforf Lectures Delivered in the University of Edinburgh.

Whitehead, Alfred North. 1934. *Nature and Life.* Cabridge University Press.

———. 1938. *Modes of Thought.* New York: Macmillan.

Wolfram, Stephen. 2002. *A New Kind of Science.* Vol. 5. Wolfram Media Champaign.

Young, GF, L Scardovi, A Cavagna, I Giardina, and NE Leonard. 2013. "Starling Flock Networks Manage Uncertainty in Consensus at Low Cost." *PLoS Comput Biology* 9 (1).

Zenil, Hector. 2010. "Compression-based Investigation of the Dynamical Properties of Cellular Automata and Other Systems." *Complex Systems* 19, no. 1 (October): 1–28.

———. 2012. "Information theory and computational thermodynamics: Lessons for biology from physics." *Information* 3 (4): 739–750.

———. 2013. *A computable universe: understanding and exploring nature as computation.* World Scientific.

———. 2017. "Reprogramming Matter, Life, and Purpose." *International Journal of Unconventional Computing.*

Zenil, Hector, N A Kiani, F Marabita, Y Deng, S Elias, A Schmidt, G Ball, and J Tegnér. 2017. "An Algorithmic Information Calculus for Causal Discovery and Reprogramming Systems."

Zenil, Hector, N A Kiani, and J Tegnér. 2017. "Low Algorithmic Complexity Entropy-deceiving Graphs." *Physical Review E* 90.

Zenil, Hector, and Elena Villarreal-Zapata. 2013. "Asymptotic behaviour and ratios of complexity in cellular automata rule spaces." *International Journal of Bifurcation and Chaos* 13, no. 9 (March).

Zipf, G. K. 1949. *Human Behavior and the Principle of Least Effort.* Martino Fine Books.

Zuboff, Arnold. 1982. "The Story of a Brain, The Mind's I." Bantam Books.

APPENDIX A

STATEMENT OF CO-AUTHOR PERMISSIONS

All co-authors have granted their permissions to use articles A. Adams et al. 2017, A. M. Adams et al. 2017, and Adams and Walker 2017 for Chapters 3, 4, and 6 respectively.

APPENDIX B

SUPPORTING INFORMATION: FORMAL DEFINITIONS OF UNBOUNDED EVOLUTION AND INNOVATION REVEAL UNIVERSAL MECHANISMS FOR OPEN-ENDED EVOLUTION IN DYNAMICAL SYSTEMS

Supporting Information: Formal Definitions of Unbounded Evolution and Innovation Reveal Universal Mechanisms for Open-Ended Evolution in Dynamical Systems

Alyssa Adams[1,2,3], **Hector Zenil**[3,4,5], **Paul C.W. Davies**[1], **and Sara Imari Walker**[1,6,7,8,*]

[1]Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe AZ USA
[2]Department of Physics, Arizona State University, Tempe AZ USA
[3]Algorithmic Nature Group, LABORES, Paris, France
[4]Department of Computer Science, University of Oxford UK
[5]Unit of Computational Medicine, SciLifeLab, Department of Medicine Solna, Karolinska Institute, Stockholm Sweden
[6]School of Earth and Space Exploration, Arizona State University, Tempe AZ USA
[7]ASU-SFI Center for Biosocial Complex Systems, Arizona State University, Tempe AZ USA
[8]Blue Marble Space Institute of Science, Seattle WA USA
[*]To whom correspondence should be addressed: sara.i.walker@asu.edu

ABSTRACT

Supporting information and techincal details for the manuscript *Formal Definitions of Unbounded Evolution and Innovation Reveal Universal Mechanisms for Open-Ended Evolution in Dynamical Systems*.

1 Description of Implementations of Time-Dependent Cellular Automata Variants

We consider three new variants of cellular automata (CA) to identify mechanism(s) that can produce conditions necessary for open-ended evolution (OEE) in bounded regions, subject to the formal criteria for OEE laid out in Definitions 1 and 2 in the main text. We consider definitions of unbounded evolution (UE) and innovation (INN) that are applicable to *any* instance of a dynamical system u that can be decomposed into two interacting subsystems o and e. Each CA variant implements *time-dependent* rules for o, with different functional forms f for this time-dependence. Here we describe in detail the implementation of each variant considered.

1.1 Case I: Deterministic State-Dependent Rules in Subsystem o

The first variant, Case I, implements *state-dependent* update rules. Case I CA are composed of two spatially separate, fixed-width, 1-dimensional CA: an "organism" o and an environment e. Both o and e are implemented with periodic boundary conditions, and utilize the alphabet $\{0,1\}$. The environment e is an execution of an ECA, and is evolved according to a fixed rule drawn from the set of 256 possible ECA rules, with periodic boundary conditions.

The subsystem o updates its rule according to a function f such that $r_o(t+1) = f(s_o(t), r_o(t), s_e(t))$, where s_o and r_o are the state and rule of the organism and s_e is the state of the environment. It is evolved with periodic boundary conditions. The expressed ECA rule of o at time t, $r_o(t)$, is represented by the eight-bit binary representation of its rule table[1], *e.g.* an o implementing Rule 30 at time t will have $r_o(t) = [0,0,0,1,1,1,1,0]$ (see main text Fig. 3). We refer to individual bits within the rule by the index i such that $r_o(t)[1] = 0$, $r_o(t)[2] = 0$, $r_o(t)[3] = 0$, $r_o(t)[4] = 1$ *etc.* for an o implementing Rule 30 at time t. The binary representation of ECA rules are structured such that each successive bit i iterated in this manner represents the output of application of the rule to the ordered set of triplet states $S^3 = [111,110,101,100,011,010,001,000]$.

The function f for our example implementation of state-dependent CA is constructed such that at each time-step t it compares the normalized frequency of each triplet i in S^3 in the state of o and e, $s_o(t)$ and $s_e(t)$, respectively, and flips the corresponding bit i in $r_o(t)$ if i is expressed in s_o and the normalized frequency of the triplet in $s_o(t)$ meets or exceeds the normalized frequency in $s_e(t)$ (where the frequency is normalized relative to the number of possible triplets in the state). That is, at each time-step t, a bit i in $r_o(t)$ will flip $0 \leftrightarrow 1$ if $n_i(s_o(t)) \geq n_i(s_e(t))$, where n_i counts the relative frequency of triplet i. Formally,

$$r_o(t+1)[i] = \begin{cases} \overline{r_o(t)[i]} & \text{if} \quad n_i(s_o(t)) \geq n_i(s_e(t)) \\ r_o(t)[i] & \text{if} \quad n_i(s_o(t)) < n_i(s_e(t)) \end{cases} \tag{1}$$

where the overbar represents logical negation.

An example implementation of this update function is shown in Fig. 4 in the main text, where an "organism" o with $w_o = 4$ is coupled to an environment e with $w_e = 6$, and $r_o(t) = [0,0,0,1,1,1,1,1,0]$. In the example, only for $i = 3$, corresponding to the triplet $\{1,0,1\}$, is $n_3(s_o(t)) \geq n_3(s_e(t))$. Therefore, $r_o(t+1)[3] = \overline{r_o(t)[3]} = \bar{0} = 1$, as shown schematically in Fig. 5 in the main text. In this example, the interaction of o and e under f changes r_o from Rule 30 at time-step t to Rule 62 at $t+1$.

1.2 Case II: Deterministic Time-Dependent Rules in Subsystem o

The second variant, Case II, is similarly composed of two spatially separate, fixed-width, 1-dimensional CA: an "organism" o and an environment e. As with Case I, both o and e are implemented with periodic boundary conditions, and utilize the alphabet $\{0,1\}$. The environment e is an execution of an ECA, and is evolved according to a fixed rule drawn from the set of 256 possible ECA rules, just as in Case I.

The key difference between Case I and Case II CA is that for Case II, the subsystem o updates its rule according to a function f such that $r_o(t+1) = f(s_e(t))$. That is, for Case II the update rule of o depends *only* on the state of the external environment s_e and is independent of the current state or rule of o (that is, o is *not* self-referential). Formally,

$$r_o(t+1)[i] = s_e(t)[i] \tag{2}$$

Here $r_o(t)$ is determined uniquely by $s_e(t)$, such that the binary representation of each possible state of the environment uniquely maps to one ECA rule according to Wolfram's binary classification scheme[1]. For this implementation the environment must be of width $w_e = 8$ to mediate a bijective map between $\{s_e\}$ and $\{r_o\}$. Case II CA emulate systems where the rules for dynamical evolution are modulated exclusively by the time evolution of an external system.

1.3 Case III: Stochastic Time-Dependent Rules in Subsystem o

The final variant, Case III, is composed of a single, fixed-width, 1-dimensional CA: the "organism" o. Like Case II, the rule evolution of Case III is driven *externally* and does not depend on s_o. However, here the external environment e is stochastic noise and not an ECA. In Case III CA, the subsystem o updates its rule according to a function f such that $r_o(t+1) = f(r_o(t), \xi)$, where ξ introduces random fluctuations in the implemented rule of o by stochastically flipping bits in r_o. Formally,

$$r_o(t+1)[i] = \begin{cases} \overline{r_o(t)[i]} & \text{if} \quad \xi < \mu \\ r_o(t)[i] & \text{if} \quad \xi \geq \mu \end{cases} \tag{3}$$

where μ is a fixed threshold for flipping between $[0,1)$, and ξ is a random number drawn from the interval $[0,1)$. This implements a diffusive-random walk through ECA rule space. Since the rule of o at time $t+1$, $r_o(t+1)$, depends on the rule at time t, $r_o(t)$, the dynamics of Case III CA are path-dependent in a similar manner to Case I (both rely on flipping bits in $r_o(t)$, where Case I do so deterministically as a function of s_o and s_e, and Case III do so stochastically).

2 Experimental Methods

The number of possible executions grows exponentially large with w_o, limiting the computational tractability of statistically rigorous sampling of the dynamics of each CA variant and of generating the set of counterfactual isolated ECA trajectories. We therefore explored small CA with $w_o = 3,4,\ldots 7$ and sampled a representative subset of all possible trajectories for each w_o (see Section 9 for examples of larger CA). We then generated statistics on the number of sampled trajectories satisfying Definitions 1 and 2 for unbounded evolution and innovation, respectively.

2.1 Case I Experiments

For Case I, w_e must be specified in addition to w_o. We consider systems with $w_e = \frac{1}{2}w_o$, w_o, $\frac{3}{2}w_o$, $2w_o$ and $\frac{5}{2}w_o$. For comparison to Case II and Case III CA, $w_e = w_o$ statistics are used. For each w_o and w_e, the initial state of o, $s_o(0)$, the initial state of e, $s_e(0)$, the initial rule of o, $r_o(0)$ and the rule of e, r_e, are drawn at random. For $r_0(0)$ and r_e, we only consider the 88 non-equivalent ECA rules, which dramatically reduces the number of possible cases, but still covers the full spectrum of complexity in initial configurations. We then permit r_o to evolve into any of the 256 possible ECA rules. We also ensure that no two cases sampled are initialized with the same tuple $\{s_o(0), s_e(0), r_o(0), r_e\}$.

Table 1. The size of the randomly sampled subspace for Case I CA for each w_o and w_e explored.

CA Variant	w_o	#u	% Explored	CA Variant	w_o	#u	% Explored
Case I: $w_e = \frac{1}{2}w_o$	3	2.1×10^{6}	1.25	Case I: $w_e = 2w_o$	3	3.36×10^{7}	6.92×10^{-2}
	4	4.19×10^{6}	1.25		4	2.68×10^{8}	1.73×10^{-2}
	5	1.68×10^{7}	0.62		5	2.15×10^{9}	4.69×10^{-3}
	6	3.36×10^{7}	0.62		6	1.72×10^{10}	3.16×10^{-4}
	7	1.34×10^{8}	0.31		7	1.37×10^{11}	7.75×10^{-5}
Case I: $w_e = w_o$	3	4.19×10^{6}	0.63	Case I: $w_e = \frac{5}{2}w_o$	3	6.71×10^{7}	3.91×10^{-2}
	4	1.68×10^{7}	0.31		4	1.074×10^{9}	4.88×10^{-2}
	5	6.71×10^{7}	0.16		5	8.59×10^{8}	1.22×10^{-3}
	6	2.68×10^{8}	7.81×10^{-2}		6	1.37×10^{11}	1.53×10^{-4}
	7	1.07×10^{9}	3.91×10^{-2}		7	1.1×10^{12}	3.81×10^{-5}
Case I: $w_e = \frac{3}{2}w_o$	3	8.34×10^{6}	0.28				
	4	6.71×10^{7}	6.92×10^{-2}				
	5	2.68×10^{8}	3.75×10^{-2}				
	6	2.15×10^{9}	2.52×10^{-3}				
	7	8.59×10^{9}	1.24×10^{-3}				

The space of all possible Case I CA executions is too large to explore the full space computationally. Since each e and o are each initiated with a state and a rule, the number of possible executions is:

$$N_U = N_R^2 \times N_{S_e} \times N_{S_o} = 88^2 \times 2^{8w_e} \times 2^{8w_o} \tag{4}$$

where N_R is the number of sampled initial rules for o and e, N_{S_e} is the number of sampled initial states for e, and N_{S_o} is the number of sampled initial states for o. For $w_o = w_e = 3$ and $w_o = w_e = 4$, exploring the full space of all possible initial conditions is computationally tractable, and verifies that the statistics reported herein for executions of o that display UE and INN are characteristic of the full computational space for the smaller sample sizes implemented in this study. The number of randomly sampled cases for Case I CA included herein is given in Table 1.

2.2 Case II Experiments

For Case II, $w_e = 8$ for all simulations, since this permits a bijective map from $\{s_e\}$ to the rule space of ECA and thus the set of rules $\{r_o\}$. As with Case I CA, executions are initialized with a randomized tuple $\{s_o(0), s_e(0), r_o(0), r_e\}$, ensuring that no two experiments are initialized with the same tuple. We restrict attention only to $w_e = 8$ for Case II experiments in this study to directly compare to our Case I and Case III CA. The number of randomly sampled cases for Case II CA is given in Table 2.

Table 2. The size of the randomly sampled subspace for Case II CA for each w_o explored.

CA Variant	w_o	#u	% Explored
Case II: $w_e = 8$	3	1.34×10^{8}	2.31×10^{-2}
	4	2.68×10^{8}	2.1×10^{-2}
	5	5.37×10^{8}	2.1×10^{-2}
	6	1.07×10^{9}	2.1×10^{-2}
	7	2.15×10^{9}	1.98×10^{-2}

2.3 Case III Experiments

For Case III, a threshold μ for stochastic flipping of the bits in the rule table of o must be set. Results for Case III are given for $\mu = 0.5$ in the main paper, such that each outcome bit in the rule table at every time step $r_o(t)$ has a 50% probability of flipping. Results for other values are reported in Section 5. Since we evolve only the subsystem o for Case III CA, executions

are initialized with a random tuple $\{s_o(0), r_o(0)\}$. We do not restrict sampled executions to unique tuples, since a different random seed is set for each execution. The number of randomly sampled cases for Case III CA is given in Table 3.

Table 3. The size of the randomly sampled subspace for Case III CA for each w_o explored.

CA Variant	w_o	#u	% Explored
Case III: Random	3	5.24×10^5	10
	4	1.05×10^6	5
	5	2.1×10^6	5
	6	4.19×10^6	5
	7	8.39×10^6	5

3 Calculating Recurrence Time, Compressibility and Lyapunov Exponent

Recurrence times for the state- and rule-trajectory of o were calculated to identify cases exhibiting UE and thus OEE. The complexity of the state trajectory $\{s_o(0), s_o(1), \ldots s_o(t_r)\}$ was measured by means of its *compressibility* (C), and calculation of the *Lyapunov exponent* (k).

3.1 Recurrence Time

For Cases I and II,we measured the recurrence times t'_r and t_r for o, for both the rule evolution $\{r_o(t_1), r_o(t_2), r_o(t_3) \ldots r_o(t'_r)\}$ and the state evolution $\{s_o(t_1), s_o(t_2), s_o(t_3) \ldots s_o(t_r)\}$, respectively. Recurrence times were calculated by determining the time t_r or t'_r when the *sequence* of states or rules of o, respectively, repeated. In general, t_r and t'_r for o are not the same as for the full system u (or as each other, such that often $t_r \neq t'_r$, see Fig. 1 in the main text). We therefore first determined when u repeated the tuple $\{s_o, s_e, r_o\}$ such that $\{s_o(t'), s_e(t'), r_o(t')\} = \{s_o(t), s_e(t), r_o(t)\}$, where $t < t'$. We then determined the t_r such that $\{s_o(t_r), s_o(t_r+1), \ldots s_o(t')\} = \{s_o(t_i), s_o(t_i+1), \ldots s_o(t_r)\}$ for $t_i < t_r$ (and likewise for t'_r with the replacement r_o for s_o). The time step t_i is identified as initiation of the attractor dynamics for o. In many cases, we find attractors that are unbounded and innovative by Definitions 1 and 2, in addition to full trajectories up to recurrence. An example illustrating the expected Poincaré time for o, t_P, its recurrence time for the state trajectory t_r and the attractor size for the full system u, t_a (up to the recurrence time t' for the full system) is shown in Fig. 1.

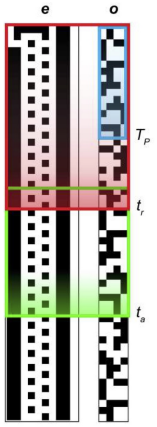


Figure 1. Relevant timescales for describing the dynamics of o embedded in u. Shown are the Poincaré recurrence time t_P (blue) for an isolated ECA of the same width w_o as o, the state-trajectory recurrence time t_r of o (red), and attractor size of the full system u, t_a (green).

Since Case III CA are stochastically evolved, their dynamics do not repeat with a unique recurrence time t_r for o. However, all executions sampled eventually terminated in an oscillation between the two homogeneous states (all-'0's or all-'1's). These states are attractors for every fixed rule ECA evolved under periodic boundary conditions, so once a Case III CA evolves to

either homogeneous state, no heterogeneity will ever be produced (the dynamics behave somewhat like dissipation of the heterogeneity in the initial state). We therefore consider it more meaningful to calculate the number of time steps before convergence to this oscillatory attractor in place of the recurrence time t_r, which we denote by t_r for consistency of notation with other cases explored. We therefore capture the timescale of relevance for all interesting (and potentially complex) dynamics, which occur in the transient before converging to this attractor.

3.2 Compressibility

The Kolmogorov-Chaitin complexity of string s is defined as the size of the shortest computer program p running on a universal Turing machine U that produces the string s (here s is the sequence of states of o):

$$K_U(s) = min\{|p|, U(p) = s\} \ . \tag{5}$$

Although it cannot be computed exactly, it is lower semi-computable and can be approximated by using a general lossless compression algorithm L[2]. This upper-bound approximation of the Kolmogorov-Chaitin complexity is normalized according to a normalized compression measure C:

$$C(s) = \frac{L(s)}{max(C_i(s), length(s))} \ . \tag{6}$$

Throughout this paper C_i is output of the Compress algorithm based on the LZW algorithm[2]. It can be replaced by the output of any other compression algorithm. The measure is therefore a family of possible indexes approximating K. We use C as measure over the state-trajectory of the organism o for each execution u, as an approximation of the characteristic complexity of o in the limit of large times $t \to \infty$.

Large values of C indicate low Kolmogorov-Chaitin complexity, meaning the output can be produced by a simple (short) program p. The normalization constant $max(C(s), length(s))$ was calculated by measuring the number of bits resulting from a generalized compression algorithm for the Poincaré recurrence time of the entire system u, not an isolated organism. This allows normalizing the observed C to its maximum possible value for an organism coupled to an environment. This closely approximates an upper limit in C for the longest possible non-repeating trajectory for any given o.

In order to ensure the normalization constant for an organism of width w_o is a close approximation to the maximal value, $C_i(s)$ was calculated for 10^7 randomly generated ECA of width w_o, evolved with a fixed rule for 2^{2w} time steps, where w is the width of u, such that $w = w_o + w_e$. The maximum of this set was used as the normalization constant $max(C_i(s), length(s))$. Thus, all C values are normalized relative to the maximal complexity of a CA evolved according to a fixed dynamical rule.

3.3 Lyapunov Exponent

The Lyapunov exponent k captures the speed at which a perturbation moves through a system[3], thereby quantifying sensitivity to initial conditions. In CA, k can be, in general, measured by perturbing a single bit in the initial condition, and counting how many bits differ compared to the unperturbed time evolution in each time step:

$$y(t) = H_{io}[s_i(t), s_o(t)] \tag{7}$$

where H_{io} is the Hamming distance between the state of the perturbed system i and the original organism o, which is evaluated at each time step t. The resulting time series of $y(t)$ values can be approximated as an exponential function, $y(t) = e^{kt}$, where k is estimated numerically. High values of k indicate sensitivity to perturbations, which is typically associated with complex dynamical systems, such as those that occur in deterministic chaos.

4 Statistics of Sampled Trajectories Displaying Innovation (INN)

Tables 4 and 5 show the resulting statistics for sampled o that were found to be innovative (INN) according to Definition 2. Table 4 includes all three CA variants as well as ECA counterfactual trajectories used as a control. Table 5 shows results for state-dependent Case I CA as a function of varying environment size w_e. Innovative o were identified as having a state-trajectory that cannot be reproduced by any closed, fixed rule ECA of equivalent width $w = w_o$.

5 Recurrence time frequency distributions

The frequency distribution of t_r observed for sampled state trajectories of o in Case I, Case II and Case III CA are shown in Figures 2,3 and 4, respectively. Comparing the three cases reveals that for equivalently sized ensembles of sampled trajectories for Case I, Case II, and Case III CA, the Case I CA generate OEE cases with higher statistical certainty than either the Case II or

Table 4. Percentage of sampled cases displaying INN for each CA variant.

w_o	**ECA**	**Case I** ($w_o = w_e$)	**Case II**	**Case III**
3	0	54.62	99.98	99.82
4	0	74.66	99.97	99.87
5	0	92.56	99.97	99.92
6	0	88.14	99.97	99.94
7	0	97.14	99.97	99.97

Table 5. Percentage of sampled cases displaying INN for Case I, with varying environment size w_e.

w_o	$w_e = \frac{1}{2}w_o$	$w_e = w_o$	$w_e = \frac{3}{2}w_o$	$w_e = 2w_o$	$w_e = \frac{5}{2}w_o$
3	30.72	54.62	70.10	86.04	93.29
4	33.32	74.66	86.57	95.52	97.47
5	32.42	92.56	96.22	98.32	98.72
6	35.64	88.14	97.03	98.91	99.29
7	52.92	97.14	97.43	99.51	99.63

Case III CA for most parameters explored. This is especially true for cases where $w_e > w_o$ in Case I simulations. From Figure 2 it is evident that larger environments yield more UE cases with $t_r > t_P$ and in general result in longer observed recurrence times.

Case II CA yield fewer OEE cases as w_o increases, as evident in Figure 3. As discussed in the main text, Case II is not scalable as it would require changing the structure of the rules of the organism o.

Case III CA generate fewer OEE cases than Case I as the width of o increases, with no cases observed in our statistical sample for $w_o > 7$ for $\mu = 0.5$ (Figure 2 bottom panel, leftmost column). The frequency distribution of recurrence times for Case III CA with $\mu = 0.01$, $\mu = 0.1$ and $\mu = 0.5$ are shown in Figure 2. For smaller values of μ OEE cases are observed for larger w_o. In the context of biological evolution, the mechanism for increasing the number of OEE cases under Case III would therefore be for systems to evolve toward *slower* mutation rates over time. However, because the distributions are exponentially distributed, the number of OEE cases is always exponentially suppressed, representing only a small tail of the distribution. A fixed width w_o execution could always be found in a *large enough* statistical sample that such that the observed t_r would be greater than the maximum recurrence time observed in a Case I CA with an equivalent organism width w_o. However, in general due to the exponential suppression of cases with larger t_r for Case III variants (Figure 4), the ensemble size of sampled cases will necessarily be much larger for Case III CA than for Case I CA. That is, for a sufficiently large ensemble size one could chose a value of μ and generate a trajectory in a width w_o organism with a given recurrence time t_r, but would always be able to find an example trajectory of the same t_r for a *smaller* sized ensemble of Case I CA for some environment width w_e. Due to the exponential suppression, OEE cases are much rarer for Case III than Case I CA. Additionally, once Case III reach the terminal attractor their dynamics are not complexity, whereas Case I CA will repeat an attractor state that is in general complex and is often times open-ended (such that the attractor itself satisfies Definitions 1 and 2). We therefore regard Case III to not be scalable.

6 ECA Rule Complexity of Case I CA

To determine if the complexity observed in Case I CA is *intrinsic* to the state-dependent mechanism, or is an artifact of a selection-effect favoring complex rules, we determined the frequency of rules implemented in Case I CA utilizing the Wolfram classification scheme for Elementary Cellular Automata[1]. There are four Wolfram Classes: Class I and II are regarded as the least complex, often generating simple repeating patterns. Class III rules are more complex displaying random patterns, and Class IV are regarded as the most complex, displaying rich dynamical structure (for example, ECA Rule 110, which is known to be Turing Universal[4] is a Class IV ECA). We analyze the complexity of ECA rules implemented in the rule trajectories of Case I CA by considering the frequency of implementation of rules from each class to determine if the complexity of the observed dynamics is an artifact of the ECA rules or *intrinsic* to f.

The resulting rank ordered frequency distribution of rules is shown in Figure 5 for all sampled Case I CA of a given organism width w_o, and separately for the OEE cases in Figure 6. Since this data includes statistics for the entire sample of o of a given width w_o included in our study, we call these distributions "metagenomes" to indicate that they represent bulk statistics over many instances of "organisms" o. The resulting distributions indicate that Case I CA primarily implement Class I and II rules, indicative that the complexity observed is *intrinsic* to the state-dependent mechanism and not an artifact of selective use of 'complex' Class III and IV ECA rules. This is true for statistics sampled over all Case I CA (Figure 5), as well as isolating only OEE cases (Figure 6).

7 Distributions of Attractor Sizes

Figures 7 - 8 show box-whisker plots of the attractor sizes for the subsystem o for each ECA and for Case I and Case II CA (Case III CA terminate in a random, oscillatory attractor and the statistics are therefore not included here, see Section 3.1 for discussion). In each figure, the black horizontal line indicates where $t_r/t_P = 1$, where t_a is the attractor size and t_P is the expected Poincaré time of an equivalent isolated system (an ECA). Sampled attractors exhibiting unbounded evolution (UE) have $t_a/t_P > 1$ and therefore fall above the black solid line - these are examples of OEE attractors.

8 Compressibility and Lyapunov Exponent Values for Case I and II CA

8.1 Compressibility and Lyapunov exponent for OEE trajectories sampled from Case I CA

Calculated values for compressibility (C) and Lyapunov exponent (k), as defined in Section 3, are shown in Figure 10 for the state trajectory of o for all sampled OEE executions for Case I CA. Comparison of the left panel of Figure 10 with the left panel of Figure 2 in the main text reveals that the observed C for all OEE cases tends to be lower than that calculated over all sampled o for Case I: that is, the OEE cases exhibit lower C, consistent with intuition that systems with longer recurrence times should be 'more complex'. As w_o increases, more OEE cases tend to have lower C values, such that larger "organisms" are more complex.

Likewise, comparing the right panel of Figure 10 with the right panel of Figure 2 in the main text indicates that OEE cases also tend to have much higher k values than that calculated over all sampled o for Case I, indicative of greater sensitivity to perturbations in OEE systems. Large values of w_o lead to larger k, on average, such that larger OEE "organisms" are more sensitive to perturbations and therefore display richer, more complex dynamics.

8.2 Comparison of compressibility and Lyapunov exponent for Case I and Case II CA

Calculated values for compressibility (C) and Lyapunov exponent (k), as defined in Section 3, are compared for Case I and Case II CA in Figures 11 and 12. Case III is not considered as the long-term dynamics display low complexity for the oscillatory attractor of the homogenous all-'0' and all-'1' states. For both Case I and Case II CA variants, increasing w_o yields more OEE cases with lower C values, such that larger "organisms" are more complex (Figure 11).

Case II CA yield lower C values than Case I for the data shown as a result of the difference in the normalization implemented in Eq. 6, which for Case II CA is lower since the the width of u is $w = w_o + 8$ for all w_o explored, whereas for Case I CA the width of u is $w = 2 * w_o$ (for $w_e = w_o$ as shown). Additionally, as noted Case I is scalable as w_e can be increased to generate higher complexity (lower C) cases. The Lyapunov exponent for Case II is in general higher than for Case I, indicating greater sensitivity to perturbations in the initial condition for Case II CA than Case I. For both CA variants, k increases with increasing organism size k, such that larger organisms are more complex.

9 Larger Systems

Figure 13 shows example executions of Case I state-dependent CA for large organisms of width $w_o = 101$, which visually demonstrate that the novelty of the dynamics reported herein scale to large system sizes.

References

1. Wolfram, S. *A New Kind of Science*, vol. 5 (Wolfram Media Champaign, 2002).

2. Zenil, H. & Villarreal-Zapata, E. Asymptotic behaviour and ratios of complexity in cellular automata rule spaces. *International Journal of Bifurcation and Chaos* **13** (2013).

3. Tisseur, P. Cellular automata and lyapunov exponents. *Nonlinearity* **13**, 1547 (2000).

4. Cook, M. Universality in elementary cellular automata. *Complex systems* **15**, 1–40 (2004).

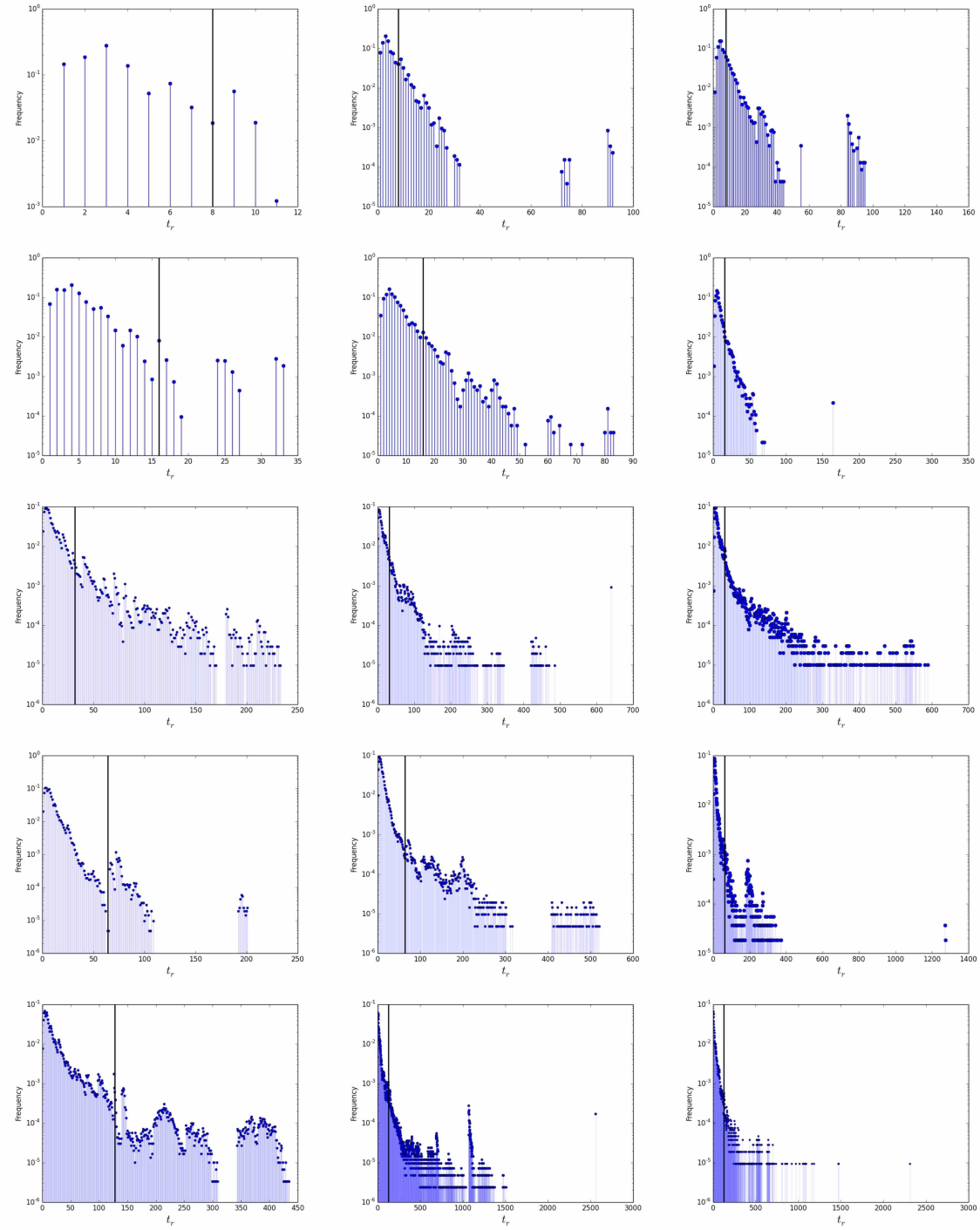


Figure 2. Frequency distributions of recurrence times t_r for Case I CA with $w_e = w_o$ (leftmost column), $w_e = \frac{3}{2}w_o$ (left middle), $w_e = 2w_o$ (right middle) and $w_e = \frac{5}{2}w_o$ (rightmost column). For rows from top to bottom, $w_o = 3, 4, 5, 6$ and 7 respectively. The Poincaré recurrence time t_P of an isolated ECA of width w_o is highlighted by the black vertical line in each panel.

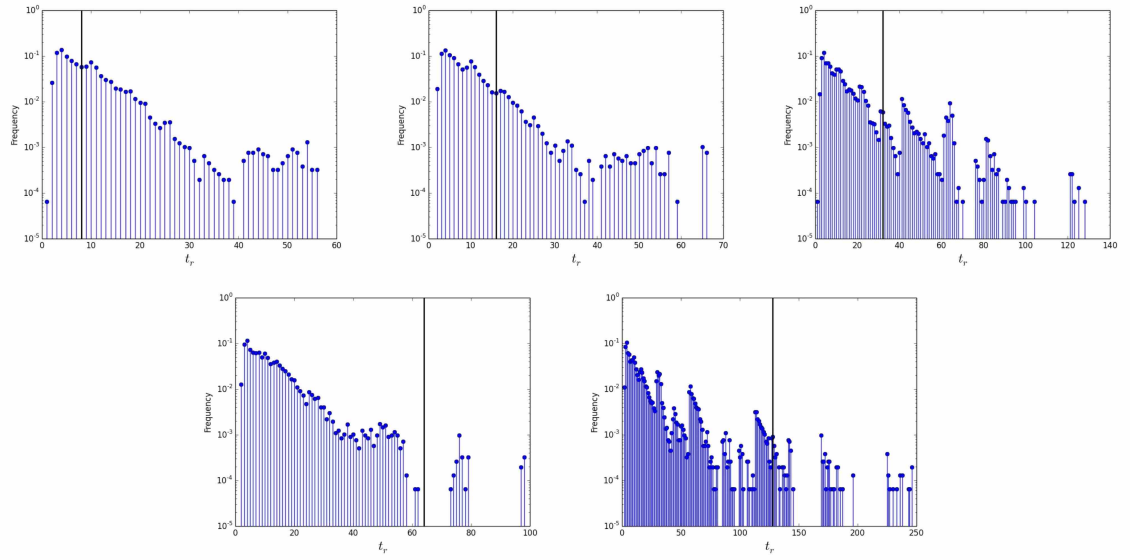


Figure 3. Frequency distributions of recurrence times t_r for Case II CA. From top to bottom, $w_o = 3,4,5,6$ and 7 respectively. The Poincaré recurrence time t_P of an isolated ECA of width w_o is highlighted by the black vertical line in each panel.

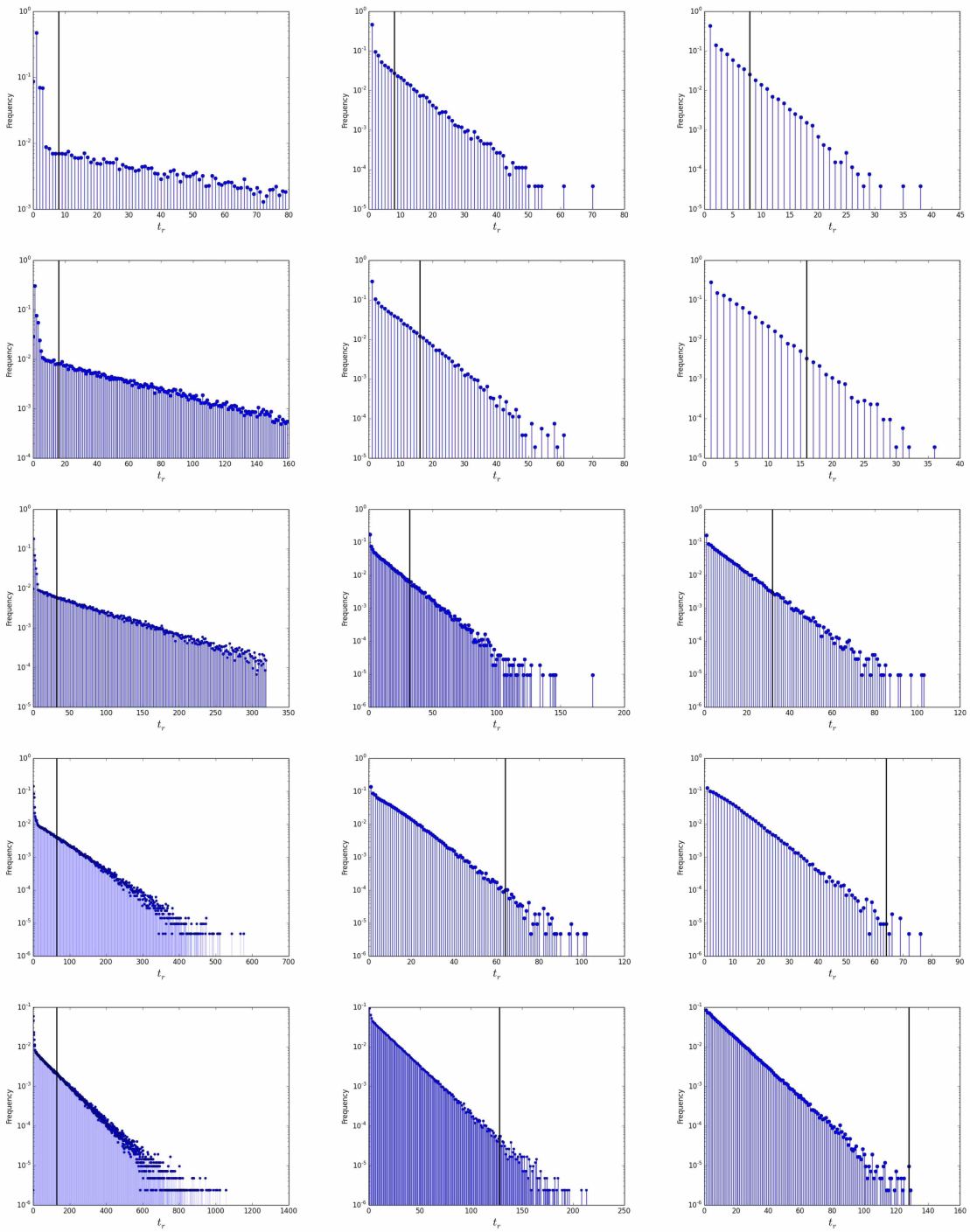


Figure 4. Frequency distributions of t_r for Case III CA with $\mu = 0.01$ (left), $\mu = 0.1$ (middle) and $\mu = 0.5$ (right). From top to bottom, $w_o = 3, 4, 5, 6$ and 7 respectively. The Poincaré recurrence time t_P of an isolated ECA of width w_o is highlighted by the black vertical line in each panel.

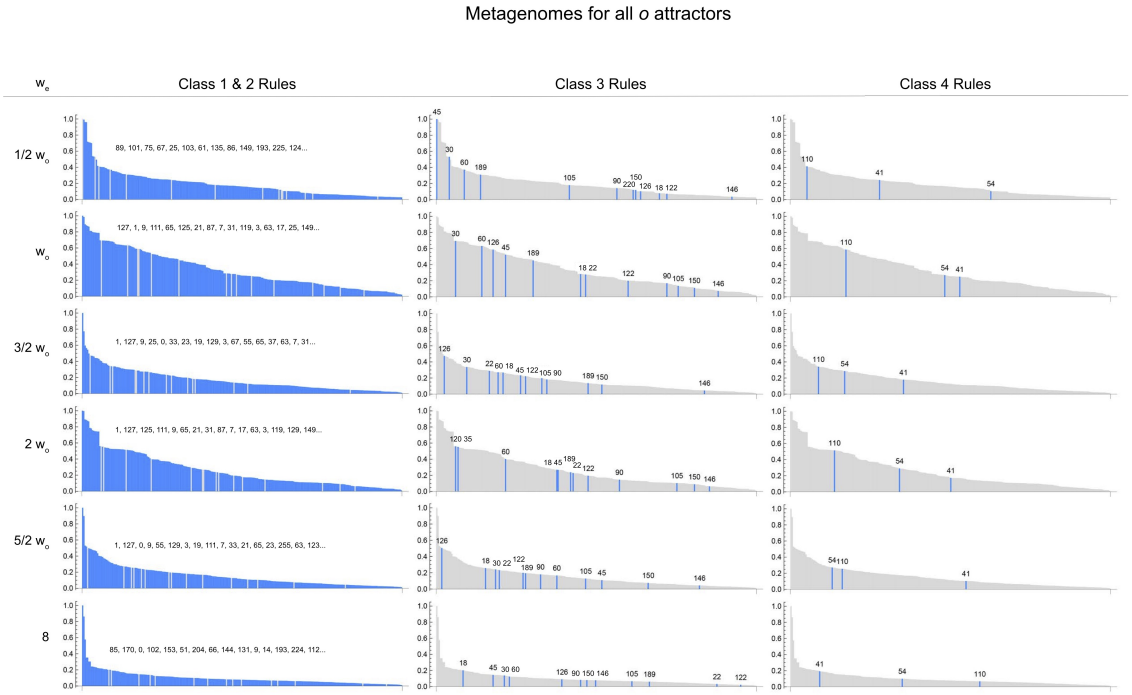


Figure 5. Rank ordered frequency distributions of rules ("metagenomes") implemented by o in its attractor. From top to bottom $w_e = \frac{1}{2}w_o$, w_o, $\frac{3}{2}w_o$, $2w_o$ and $\frac{5}{2}w_o$, respectively. Highlighted in blue are the frequencies of Class I and II rules (left), Class III rules (middle) and Class IV rules (right).

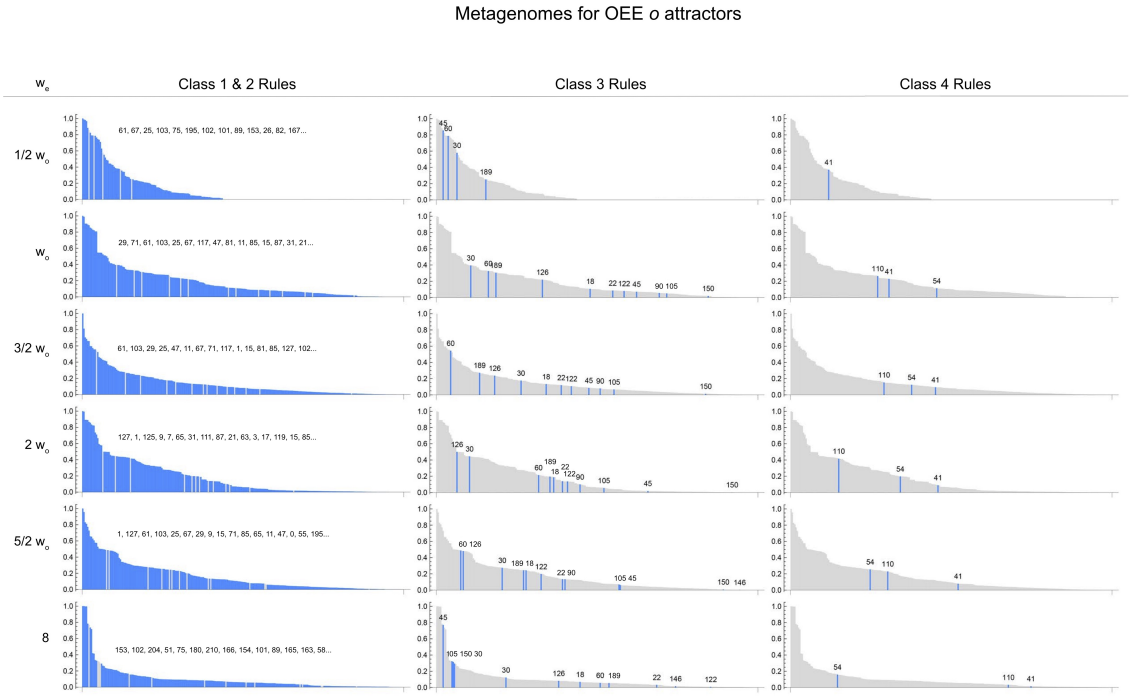


Figure 6. Rank ordered frequency distributions of rules ("metagenomes") implemented by o in its attractor for OEE cases only. From top to bottom $w_e = \frac{1}{2}w_o$, w_o, $\frac{3}{2}w_o$, $2w_o$ and $\frac{5}{2}w_o$, respectively. Highlighted in blue are the frequencies of Class I and II rules (left), Class III rules (middle) and Class IV rules (right).

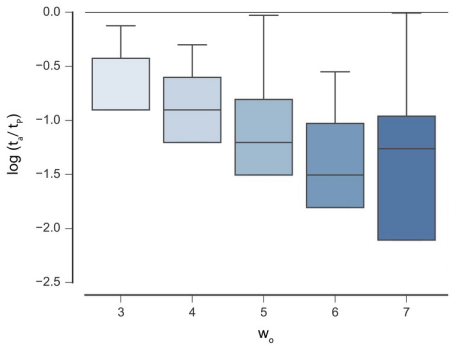


Figure 7. Distribution of attractor sizes t_a for the state trajectory of for all 88 non-equivalent ECA rules, evolved from all possible initial conditions of width w_o. Attractor sizes are normalized to the Poincaré time $t_P = 2^{w_o}$ for an isolated ECA, where the black horizontal line indicates where $t_r/t_P = 1$ (shown on a log scale).

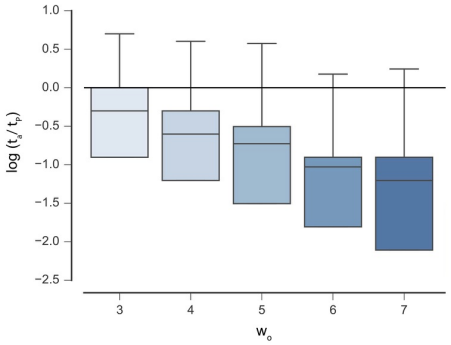


Figure 8. Distribution of attractor sizes t_a for the state trajectory of o, for Case II CA. Attractor sizes are normalized to the Poincaré time $t_P = 2^{w_o}$ for an isolated ECA. The black horizontal line indicates where $t_a/t_P = 1$ (shown on a log scale). Sample trajectories displaying *unbounded evolution* (UE) occur for $t_a/t_P > 1$.

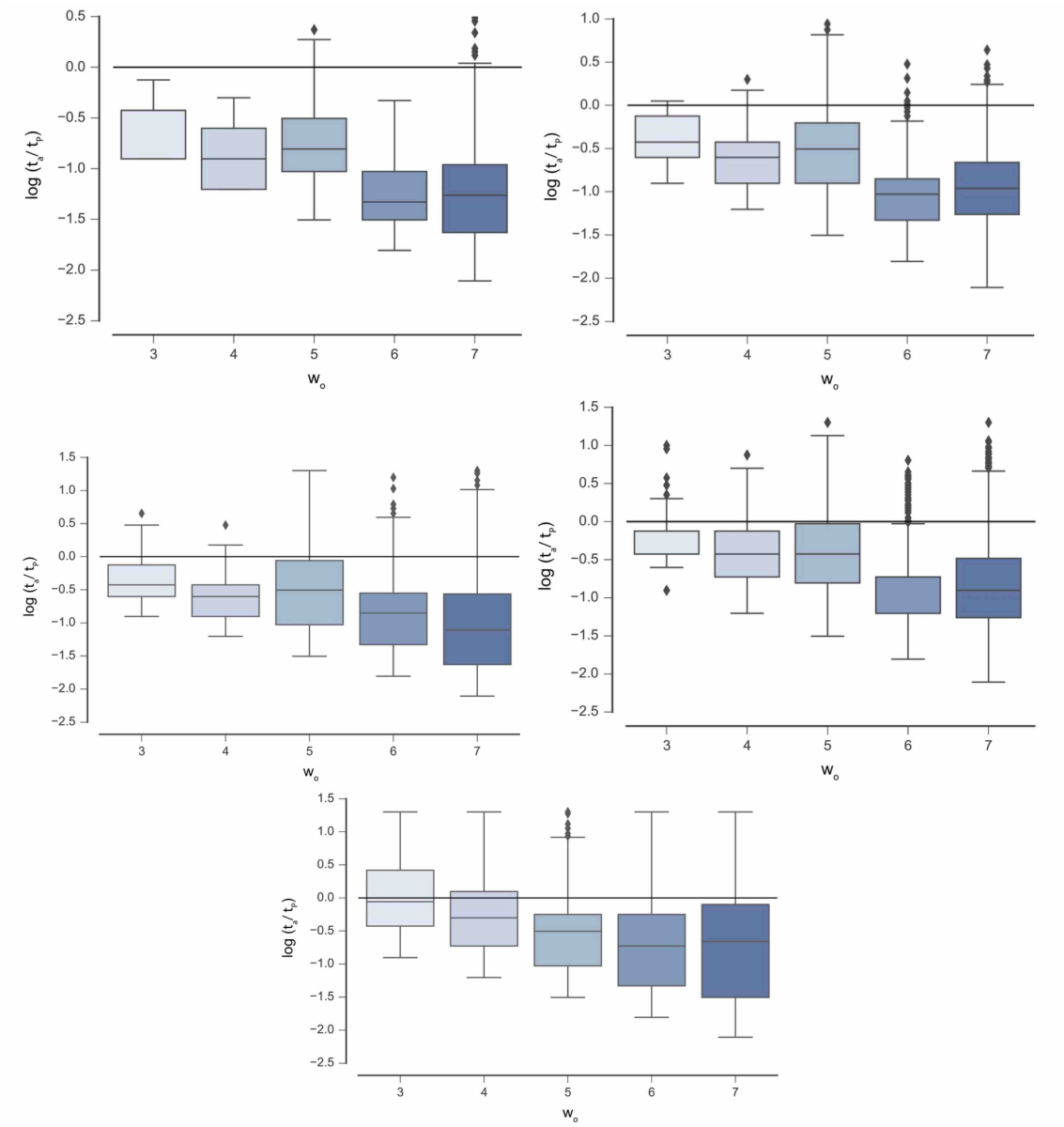


Figure 9. Distribution of attractor sizes t_a for the state trajectory of o for Case I CA. Shown from top to bottom are distributions for $w_e = \frac{1}{2}w_o$, w_o, $\frac{3}{2}w_o$, $2w_o$ and $\frac{5}{2}w_o$, respectively. Attractor sizes are normalized to the Poincaré time $t_P = 2^{w_o}$ for an isolated ECA. The black horizontal line indicates where $t_a/t_P = 1$ (shown on a log scale). Sampled trajectories displaying UE occur for $t_r/t_P > 1$.

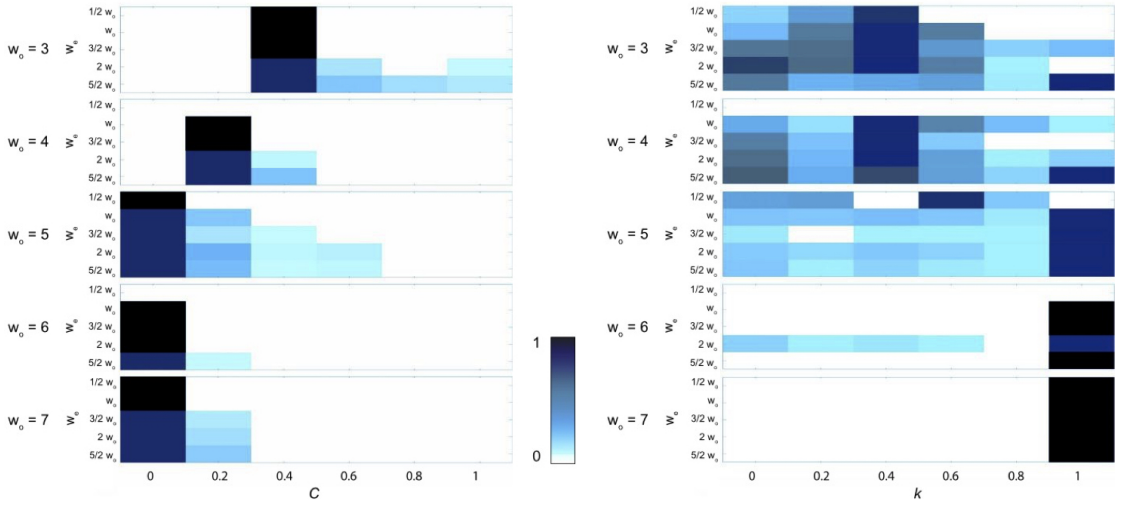


Figure 10. Heat maps of compression C (left) and Lyapunov exponent values k (right) for sampled OEE trajectories for the states of o for Case I CA. From top to bottom $w_o = 3,4,5,6$ and 7, with distributions shown for $w_e = \frac{1}{2}w_o$, w_o, $\frac{3}{2}w_o$, $2w_o$ and $\frac{5}{2}w_o$ (from top to bottom, respectively) for each w_o. Distributions are normalized to the total size of sampled trajectories for each w_o and w_e (see statistics in Table 1).

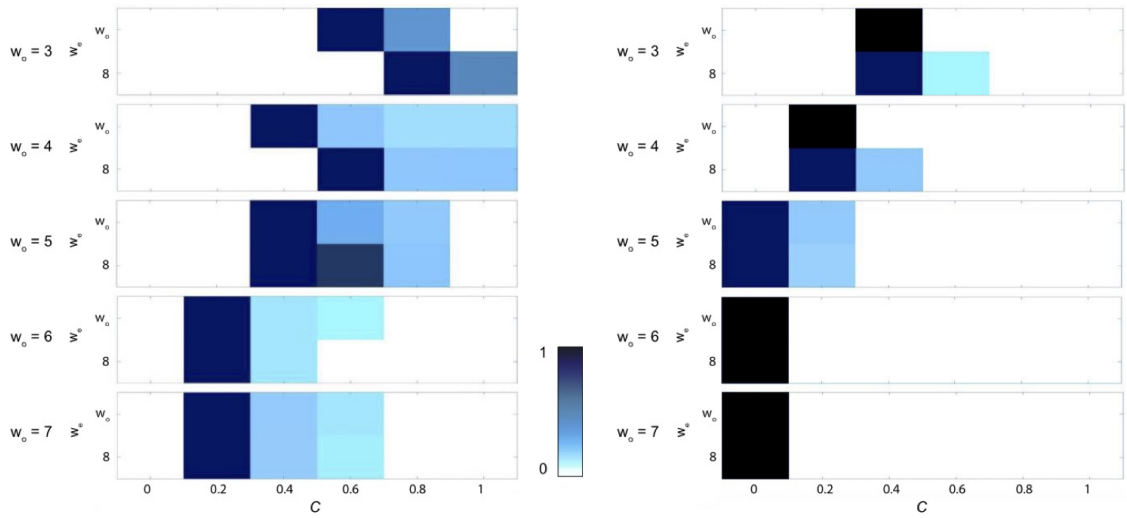


Figure 11. Heat maps of compression C for all sampled trajectories of the states of o (left), and for OEE trajectories only (right) shown for Case I and Case II CA. From top to bottom $w_o = 3,4,5,6$ and 7. For each w_o shown are the distributions of C for Case I CA for $w_e = w_o$ (top row in each panel) and for Case II CA with $w_e = 8$ (bottom row in each panel). Distributions are normalized to the total size of sampled trajectories for each w_o and w_e for each CA variant (see statistics in Tables 1 and 2).

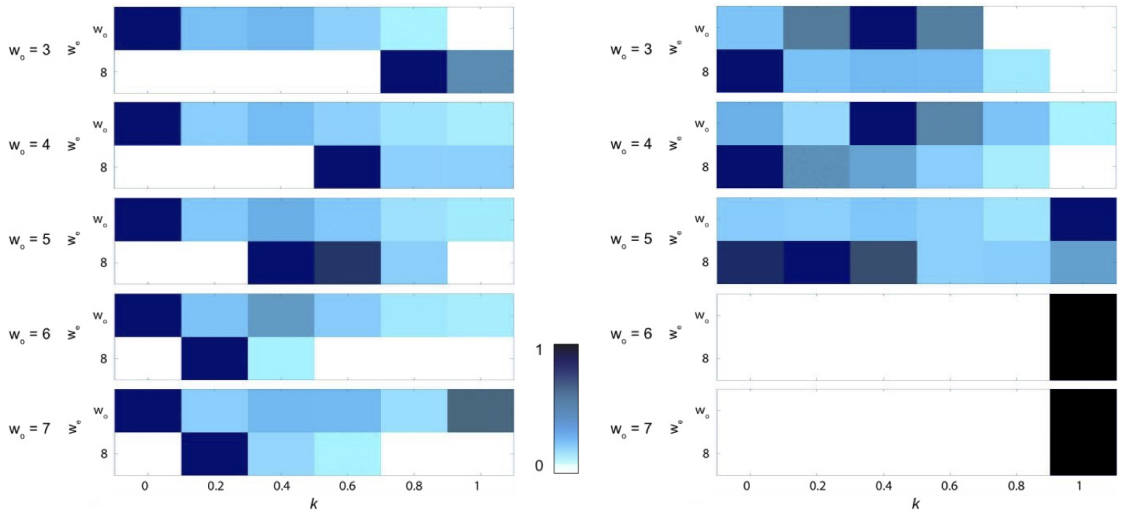


Figure 12. Heat maps of Lyapunov exponent k for all sampled trajectories of the states of o (left), and for OEE trajectories only (right) shown for Case I and Case II CA. From top to bottom $w_o = 3, 4, 5, 6$ and 7. For each w_o shown are the distributions of k for Case I CA for $w_e = w_o$ (top row in each panel) and for Case II CA with $w_e = 8$ (bottom row in each panel). Distributions are normalized to the total size of sampled trajectories for each w_o and w_e for each CA variant (see statistics in Tables 1 and 2).

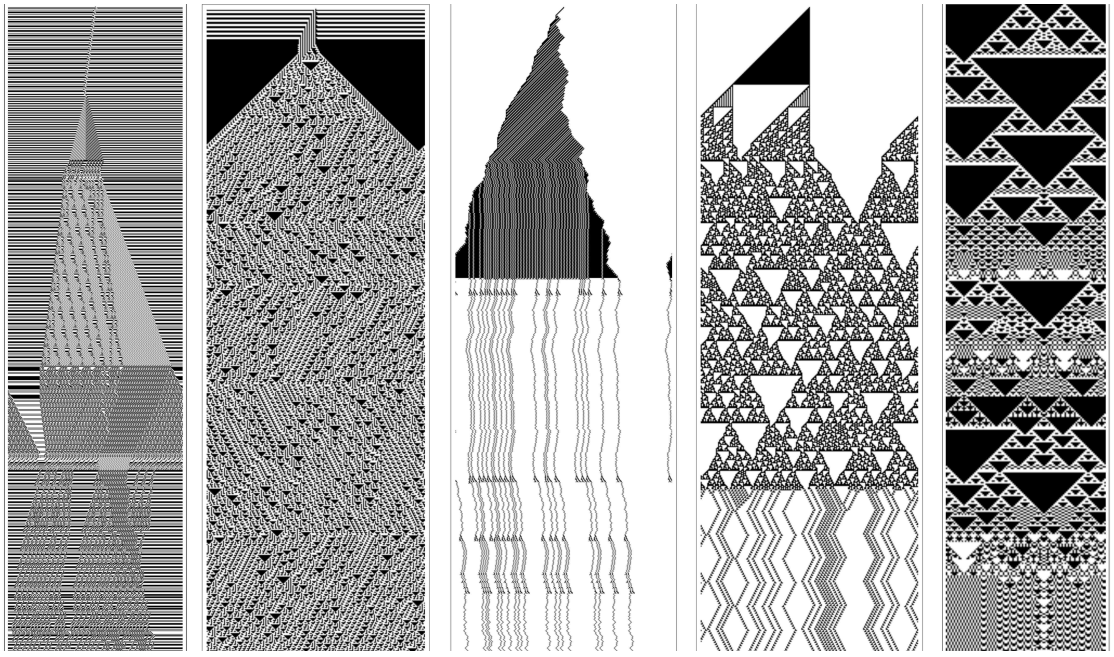


Figure 13. Example executions of the state trajectory of o for Case I CA for large system size $w_o = 101$, with $w_e = w_o$.