Asking Big Questions: What is a physics of living systems?[1]

Pankaj Mehta

January 30, 2019

[1] Readings: "Von Hixton Lecture", Von Neumann; "What is Life?", Schrödinger. I would also like to thank Hyman Hartman for inspirational discussions that led to this framing.

In this lecture, we ask the "big questions": What is life? What should a physics of living systems look like? We take inspiration from three old theorists: Aristotle, Darwin, Von Neumann who offered three crucial insights: life can be empirically studied and classified with form following function, life is historical, and the self-replication that defines life is related to "computation". We then discuss Schrödinger's "What is Life?" which inspired many of the founders of Molecular Biology and sought to find the answers in molecules - or aperiodic crystals as Schrödinger called them. We then discuss why the reductionism of molecular biology, though extremely powerful, is not sufficient for understanding life. In particular, any understanding of life must confront four major challenges: self-organization and the emergence of complexity, the nonequilibrium basis of life, its historical nature, and its ability to transmit information with high-fidelity (computational aspects) . We end the lecture by discussing Von Neumann's beautiful thought experiment relating self-reproduction in life to Turing's theory of universal computation. We point out what Sydney Brenner called Schrödinger's Fatal Error and how Von Neumann fixed this. We conclude by introducing the central dogma (DNA, RNA, Proteins) and discuss how this is a particular realization of the Von Neumann schema.

One of the things we seldom do is ask big questions. Here we will ask big questions: What does it mean to be alive? If we went to another solar system, how would we know that we encountered life? How is a physics of living systems different from the physics of inanimate matter? Before going on, I think its worth taking a few minutes to brain storm answers to this question. I encourage everyone to spend 5-10 minutes thinking about this and ideally discussing this with someone. [2]

[2] In class, divide into small groups, then discuss as a class. Make list on the board to try to really tackle the difficulty of this question.

Three great theorists of life

According to Hyman Hartman, there have only three great theorists of life: Aristotle, Darwin, and Von Neumann. What are their contributions:

- **Aristotle:** Life is not random. It can be studied empirically and classified. Furthermore form follows function.
- **Darwin:** Life is historical and shaped through evolution. The complexity we see in modern life is "evolved".

- **Von Neumann:** Life is intimately related to computation and information.

Aperiodic crystals and the hegemony of Molecular Biology

One of the most influential contributions of theoretical physics to biology is a short pamphlet written by Schrödinger in 1944 called "What is Life?". This book inspired countless seminar figures in early molecular biology such as Sydney Brenner, Max Delbruck, and Francis Crick. Let us briefly think about what the book says.

Why should we be surprised by life at the scales is exists?

- *Molecules exists at microscopic scales*. Here we know that Brownian motion dominates. This was insight due to Einstein. What is Brownian motion? It is essentially the idea that thermal noise matters. Why does it not matter for macroscopic quantities. Well we know that the thermal diffusion constant D scales as

$$D \sim \frac{k_B T}{\eta a}, \tag{1}$$

  where η is the viscosity, T is the temperature, and a is the radius. How can you reproduce reliably when this dominates? [3] We will derive the laws of Brownian motion in the next lecture.

- *Precision of replication* Life seems to be very precise. Roughly speaking, we expect the error in particles to scale like $1/\sqrt{N}$. However, even at the time of Schrödinger it seemed like single particles (genes) were controlling information and replication. How could you have high quality reproduction?

- *How can you encode so much high-fidelity information in a single molecule?* The molecule must also function as a code and have some way of encoding information.

[3] Exercise: Estimate this number for a 1nm particle, 1 μ m particle, and 1 cm particle in water at room temperature? When does Brownian motion matter based on this exercise?

Schrödinger's answer: we must have "aperiodic crystal". However, there were still fundamental puzzles to Schrödinger. How do we reconcile this ability to reproduce with high-fidelity with thermodynamics? In fact, he argued that this is the key to unlocking new physics

> What I wish to make clear in this last chapter is , in short, that from all we have learnt about the structure of living matter, we must be prepared to find it working in a manner that cannot be reduced to the ordinary laws of physics. And not of the ground that there is a 'new force' ... but because the construction is different from anything we have yet tested in the physical laboratory. (Chapter 7)

This is still the real promise of biological physics.

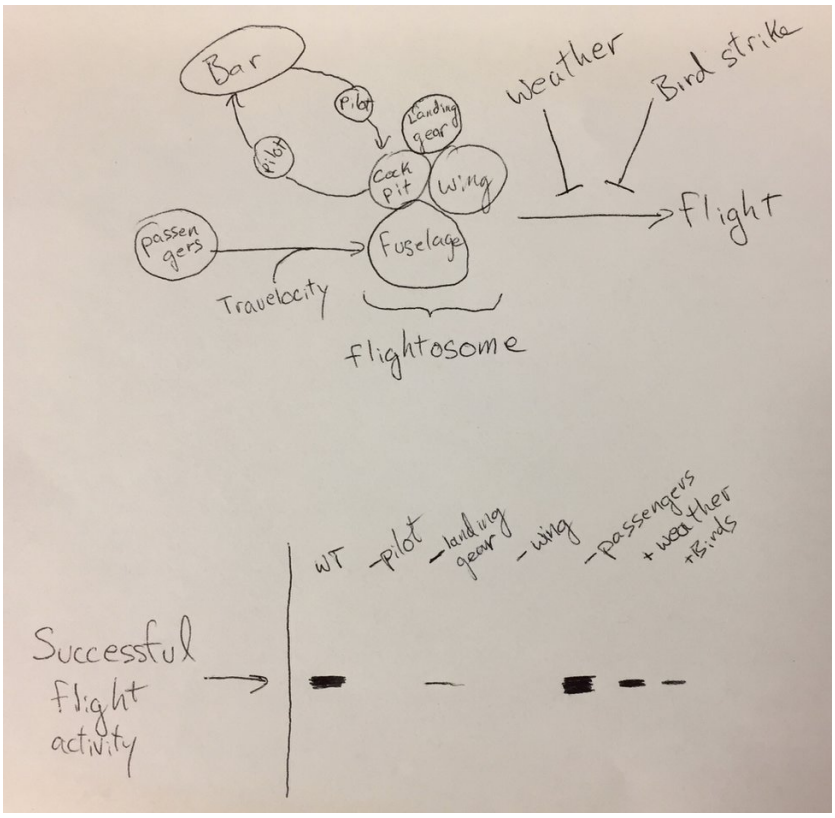

Figure 1: Molecular biology "mechanistic model" of airplane. Arjun Raj https://twitter.com/arjunrajlab/status/822202835781685252

Molecular Biology Revolution and the retreat of biological physics

We, now in fact, know that such an aperiodic structure does exist: DNA. Since its discovery, DNA has revolutionized the field of biology. In fact, when one refers to biology, almost always one now means molecular biology. In fact, "understanding" a phenomena in biology is now synonymous with finding the molecular components. The larger questions that biological physics perspective raises have disappeared. Instead, we now have parts list with interaction arrows. While this is certainly informative, it's very far from the holistic picture we should strive for. After all, this is not how we think about inanimate materials. This is captured beautifully by this satirical figure from Arjun Raj (see Fig. 1).

This biological physics tradition which actually birth molecular biology has been subsumed by the obsession with molecules. However, it is now clear that both biology and physics are ripe for a return of these questions and ideas. We are drowning in a mountain of data (sequencing, omics, microscopy, etc) and yet have no principles to organize this. The time is ripe for a physics of living systems.

Challenges and common features that define a physics of living systems

However, the questions Schrödinger raised have not disappeared. Any attempt at understanding living systems must deal with numerous challenges. I highlight a few here that I feel distinguish living systems from inanimate matter.

- Living systems operate out of equilibrium and seem to create order by consuming energy.
- Living systems seem self-organize into complex systems with heterogeneous interacting parts.
- Living systems are historical and have evolved.
- Living systems must perform complex computations and information processing tasks, responding and adapting.
- Living systems self-replicate with high precision and pass on information with high fidelity despite being made by noisy parts.

Any biological system must deal with all these complexities.

An exercise in detecting life: What can we measure and observe?

The question still remains how we can operationalize these characteristics of life. This is not just a thought exercise. For example, people from NASA

Von Neumann's argument relating life to computation

Of the four properties discussed in the last section, the relationship with computation is, perhaps, the most abstract and hardest to understand [4].

The key insight in this regard is provided by Von Neumann who related the ability of living systems to self-reproduction to the ideas of universal computation developed by Turing. Using this insight, Von Neumann essentially fixed what Sydney Brenner calls Schrödinger's Fundamental Error:

> "Schrödinger says that the chromosomes contain the information to specify the future organism and *the means to execute it* and thats not true. The chromosomes contain the information to specify the future organization and *a description of the future means to implement it, but not the means themselves.*" [5].

[4] In class, I assign students to read Von Neumann "The General and Logical Theory of Automata", especially the section entitled THE CONCEPT OF COMPLICATION: SELF- REPRODUCTION p312-319 and we have a discussion

[5] This YouTube video by Sydney Brenner is well worth watching for a discussion of these ideas `https://www.youtube.com/watch?v=5IctxzlXCiY`

This crucial distinction explains why even after we know the sequence of an organism, we still cannot really know its function. All we have is a description but not a machine to implement it. This is why its not surprising why the Human Genome Project failed to live up to all the hype that promised we would understand everything about humans once we knew our sequence [6]. This was just a description, not the machine (the cell) to read out this description and make a complicated form. In fact, the DNA sequence is only one of three crucial components that Von Neumann laid out must be part of any self-replicating machine. These four components:

[6] For an insightful critique of the Human Genome Project, see Richard Lewontin's essays collected in *It Ain't Necessarily So: The Dream of the Human Genome and Other Illusions*

- An instruction I_X that encodes for an automata X
- An constructor automata A that can takes an instruction $I_X$and execute it to create automata X
- An copyier automata B that can read an instruction I_X and copy it.
- A control automata C that interacts with A and B and causes A to construct the automata X to specify and B to copy I_X

Von Neumann's construction in greater detail

Von Neumann points out that we need two fundamental things for "self-reproducing machines" :

1. a method of copying the machine
2. a method for copying the instructions of the machine.

In this way, we can avoid the problem of what Von Neummann calls complication:

> When an automaton performs certain operations, they must be expected to be a lower degree of complication than the automaton itself. In particular, if an automaton has the ability to construct another one, there must be a decrease in complication as we go from construct to parent.

This problem is really one of the mysteries of life. How can you keep producing copies of yourself that are faithful enough to self-replicate, and even evolve to have increased complexity? Von Neumann actually shows there is an out, one that fixes Schrödinger's Fundamental Mistake.

The basic idea is to combine the basic parts mentioned above. The idea is to make a new automata that combines the basic ingredients above into a new automaton

$$D = A + B + C. \tag{2}$$

Notice that this automata will not function without an instruction I as input. The key to creating a self-replicating machine is to give automata D the instructions I_D to D. This new automata

$$E = D + I_D = A + B + C + I_D. \tag{3}$$

is our self-replicating automata. These is basic computational logic principles that must underlie life!

What is still unclear even after von Neumann's astonishing theoretical feat is how to combine this basic logic with nonequilibrium physics, self-organization, and the emergence of complexity. It seems like a full theory should be able to make "phase diagrams" of what is possible in the speed, energy, accurracy, complexity plane. Any full "theory of life" will have to contain at least these minimum requirements and is yet to be constructed. This (in my opinion) is the ultimate task of a theoretical biological physics.

The Central Dogma: a realization of the Von Neumann scheme

So how is Von Neumann's schema for a universal replicator manifested in our understanding molecular biology. This is summarized in Table 1. As can be seen, its clear that we understand the general principle that must underlie self-replication including the Central Dogma (and its necessary modifications).

At this point if you are unfamiliar with the basic biochemistry and molecular biology (DNA, RNA, proteins, ribosomes, transcription, translation, replication) you should read the accompanying crash course/refresher on how this works [7]. We will not discuss this in detail but assume that you know this basic biology.

[7] This is Chapter 1 of the wonderful forthcoming book by Anirvan Sengupa " Modeling Molecular Networks" in which I have also played a very minor role in revising

Operationalizing our understanding of life

If we had to find life, what do we look for: "hydrothermal vents", organic acids, and water. How does this fit into our discussion above of defining life? **[EXPAND]**

	Universal Replicator	Prokaryote Realization (*E. coli*)
I	Instructions for an automata	DNA
A	Automata that can read instructions I and create specified automata	transcription+translation machinery (DNA, RNA, RNA Polymerase, ribsomes)
B	Automata that can copy instructions	DNA Polymerase + specialized replication machinery (topoisomerases, condensins for segragation, etc.)
C	Control Module that initiates A+B	Gene regulation, signal transduction, etc.
D=A+B+C	Analogue of Universal Turing Machine	"Cellular machinery " minus DNA
E=D+I	Self Replicator	Self-replicating cell (i.e. *E. coli*

Table 1: The central dogma and modern molecular biology as a particular instantiation of Von Neumann's universal replicator.

Homework

1. Estimate the diffusion constant of a DNA molecule in water, the cytosol, and the nucleus. How far will they typically diffuse in 1sec, 1 minute, 1 hour? [8]. Compare this to the diffusion of a 1cm bead in water, honey.

2. Discuss how RNA viruses fit into the Von Neumann schema. Identify A,B,C,D, E for these viruses. Why does this complicate defining life?

3. Von Neumann's ideas gave rise to a lot of interest in understanding more abstract versions of life. One interesting direction this took was Cellular Automata. These automata take an initial condition and can evolve. The most famous of these is the Game of Life invented by Mathematician John Conway [9]. In the Game of Life, there are cells that live on a two-dimensional grid. Each cell can be in one of two states (alive or dead). Each cell interacts with its

[8] A useful site to know about is BioNumbers `https://bionumbers.hms.harvard.edu/search.aspx` and Phillips+ Milo's new book *Cell Biology by the Numbers*. You will have to work harder to learn about the nucleus (i.e. use Google.)

[9] A good introduction are the Wikipedia entries on Cellular Automata and Game of Life

eight neighbors (horizontal, vertical, diagonally adjacent). At each time step we have the following rules:

- Any live cell with fewer than two live neighbors dies, as if by underpopulation.
- Any live cell with two or three live neighbors lives on to the next generation
- Any live cell with more than three live neighbors dies, as if by overpopulation.
- Any dead cell with exactly three live neighbors becomes a live cell, as if by reproduction.

A simple and elegant implementation of the Game of Life is available at Jake VanderPlas' Pythonic Perambulations blog `https://jakevdp.github.io/blog/2013/08/07/conways-game-of-life/`. At the end of the entry at the bottom, you can download a Jupyter Notebook (see `https://jupyter.org/` for how to install and run). Note that you will have to make some slight changes (see instructions from Brian Dawes here):

```
For the imports, I just commented out:
    from JSAnimation.IPython_display import display_animation, anim_to_html

    And put in:
    from IPython.display import HTML

    At the bottom of the life_animation function I replaced:
    return display_animation(anim, default_mode=mode)

    with:
    return HTML(anim.to_jshtml())


Also I had some issues with the setting of figsize and dpi inside the
life_animation function. I ended up just hardcoding in a figsize of (4,3)
instead of trying to program something to calculate it.
```

- Please play around with the notebook and try 5 new initial seeds that are not mentioned in the notebook. Document what you see and discuss.

- Now modify the code so that cells with two, three, or four neighbors live. Try some of the patterns in the initial notebook. How do the outcomes differ when we change the rules?
- Discuss what this exercise tells us (if anything) about self-replication and complexity.