Why are biological systems so messy, and how can mathematicians cope?

Winfried Just
Department of Mathematics, Ohio University

April 4, 2017

Abstract

This is a leaked transcript of a conversation about biological systems, challenges for mathematical modeling of them, and insights that may result from such modeling. It highlights some work that the presenter has done in the past, some that he is trying to do at the present, and some that he would like to do in the future, if only (redacted).[1]

Ohio University – Since 1804

Department of Mathematics

[1] Several anonymous sources claim that the redacted part originally read "he could find the time;" other sources maintain that it read "he could get a student interested in them." Neither version could be independently verified.

Saturday night. A dimly lit bar on Court Street. Two patrons only; the crowds to arrive soon.

Bob: Hi. I'm Bob.

Emily: Hi. I'm Emily. What do you study?

Bob: Mathematics.

(Bracing himself for the inevitable "I'm terrible at math!")

Emily: Wonderful!! I'm a biologist and very much into math modeling. (Totally stunning radiant smile).

Bob: (Stunned.)

Emily: We are, like, kindred spirits!

Bob: (Never a fan of wasting time.)
So, as a biologist, what do you think of when you look at me?

Emily: Genes. That turn each other on

Bob: Aah! (Big, but not exactly stunning smile)

Emily: and off.

Bob: Oh!

Gene regulatory networks R us

Emily: You see, biologically speaking, deep down, who we are at any given moment, is basically the giant networks of biochemical reactions that go on at that moment in each of our cells.

Bob: But you said "genes."
I thought they just sit in the DNA at all times?

Emily: You are right. But some genes code enzymes that are needed for certain reactions to proceed, and these enzymes are only present when their genes are expressed.

Bob: And how does the cell know when to express which gene?

Emily: Other genes code transcription factors. When their concentration is sufficiently high, they will bind to certain places in the DNA and enhance or inhibit the expression of certain genes. So the vector of concentrations of these gene products, enzymes or transcription factors, will determine which genes are expressed, and thus what else will go on biochemically in a cell, and, at a higher level of biological organization, in an organism.

ODE models of gene regulation

Bob: Is this what they call gene regulatory networks?

Emily: Yes. This is what I study as a biologist.

How would you, as a mathematician, model them?

Bob: You said "concentrations." We could perhaps treat these concentrations as variables in a model based on so-called ODEs, that is ordinary differential equations.

Emily: Yes, I have worked with such models.

Bob: (Realizes that she **is** a kindred spirit.)

But hold on. If these variables change continuously, how could you say that genes turn each other off?

Emily: Or on, for that matter.

Yeah, that's a puzzler.

Boolean network models of gene regulation

Reka: (Joins them.) Simple. For each gene product, distinguish between just two concentrations: High/On (give it a value 1), and Low/Off (give it a value 0).

Then assume time proceeds in discrete steps $\tau = 0, 1, 2, \ldots$ rather than continuously.

Emily: I like this! In my lab, we take measurements of concentrations only once every hour and concentration levels of gene products in a cell are difficult to measure precisely anyway!

Reka: The values $0, 1$ are called Boolean values, and the (Boolean) concentration $s_i(\tau + 1)$ of gene product number i at time $\tau + 1$ could be modeled as a Boolean function of the concentrations $s_{j_1}(\tau), s_{j_2}(\tau), \ldots, s_{j_k}(\tau)$ at time τ of the relevant transcription factors $j_1, \ldots, j_k$.

Bob and Emily: What does "Boolean function" mean?

More on Boolean network models of gene regulation

Reka: For example, assume gene 1 has only two transcription factors, an enhancer that must be at high concentration for gene 1 to be expressed, coded by gene 2, and an inhibitor that must be at low concentration for gene 1 to be expressed, coded by gene 3.

Then $s_1(\tau+1) = s_2(\tau) \wedge \neg s_3(\tau)$.

Emily: And now, if, for example, gene 1 codes the only transcription factor that regulates the expression of genes 2 and 3 and is an enhancer for both,

Reka: then we have a Boolean network with

$s_2(\tau+1) = s_3(\tau+1) = s_1(\tau)$.

Bob: If we start with gene product 2 at high concentration and gene products 1 and 3 at low concentrations,

Reka: in state $\vec{s}(0) = (0,1,0)$,

Bob: then $\vec{s}(1) = (1,0,0)$, so that gene 1 gets turned on

Emily: and gene 2 gets turned off. Cool!

Trajectories in Boolean networks

Bob: Since $\vec{s}(1) = (1, 0, 0)$, we will now have

$\vec{s}(2) = (0, 1, 1)$, $\quad \vec{s}(3) = (0, 0, 0)$,

and $\vec{s}(t) = (0, 0, 0)$ for all $t \geq 3$.

Reka: The sequence $(\vec{s}(0), \vec{s}(1), \vec{s}(2), \dots)$ is called the trajectory of the initial state $\vec{s}(0)$.
In our example, the trajectory has reached a steady state after three time steps.

Emily: Do all trajectories in Boolean networks eventually reach a steady state?

Reka: In most Boolean networks, no. However, after perhaps going through some transient states, they always reach a periodic attractor of states that are visited, in a fixed order, infinitely often.

Bob: Cool stuff!

Emily: But ...

Do Boolean models work?

Reka: But?

Emily: These Boolean models are too simple!!

They cannot possibly make biologically realistic predictions!!!

Reka: (Calmly.) Some of them do.

I have a number of papers that demonstrate this.

Bob: What do you mean by "Do make realistic predictions?"

Concentrations take real values and change continuously, so how can your discrete-time Boolean model make correct predictions?

Reka: They predict exactly the same the sequences of sets of genes being turned on and off that have been empirically observed in the lab.

Can one prove this?

Bob: Empirically, maybe. But I am thinking mathematically. Let's say we have an ODE model as we discussed earlier. Wouldn't such a model be considered biologically more realistic, Emily?

Emily: I would think so.

Bob: Now I can think of such a model predicting a sequence of turning genes on or off based on partitioning, or discretizing real-valued concentrations into low or high ones. Or, in Reka's words, predicting Boolean trajectories. Can one actually **prove** mathematically that some such ODE model does predict **the same** Boolean trajectories as one of Reka's Boolean network models? For all time steps?

Eve: Yes. I did this for one ODE model of a small gene network.

Why should this work?

Eve: I still don't understand though whether there is an important biological reason for such an exact correspondence between ODE and Boolean models.

Emily: Maybe there is.
Think about your favorite biological function.

Bob: (Thinking about his favorite biological function.)

Emily: It may require a reliable switch from one state to another in response to a certain signal. The signal may sometimes be stronger, sometimes weaker, or even ambiguous. So the biological system must work in such a way that the switch to the response does not depend much on this noise in the signal. The response must be robust against the noise and often work more or less like a Boolean on/off switch.

The others: Yes, this sounds like a plausible reason why Boolean models often work so well as they apparently do.

Why study small networks?

Bob: But hold on. Aren't there tens of thousands of genes? How can we learn anything biologically meaningful from studying small gene regulatory networks with only a few genes?

Reka: Very large networks are analytically intractable. Fortunately, biological networks tend to have a modular architecture. This means that typically only a few genes are important for a given biological function. You can gain insights by focusing in your models on these few genes.

Emily: Not so fast. Biologists who study actual gene regulatory and other biochemical networks often report only a few chemical species that play a role, but then subsequently discover more and more of them that are involved.

Eve: Yes, and sometimes we need to take more variables into account than just a few gene products. In my model we need also the messenger RNA's that are intermediaries between genes and gene products.

Reka: I said "important," not "involved." Often we observe that when we extend a model beyond the important genes and consider additional ones that are merely involved, we get the same predictions of the Boolean trajectories for the important genes.

Why are biological systems so messy?

Bob: Are you saying that biological systems are more messy than they apparently need to be?

Reka: You can put it this way.

Bob: But why??? Why do we need all these "intermediaries" and other genes that are merely "involved"??

Emily: Perhaps there are important chemical constraints why we need messenger RNA to mediate between DNA and gene products.

Theo: Hi kids!
Nothing in biology makes sense except in the light of evolution. Biological systems are evolved, not designed.[2] Evolution is a tinkerer, not an engineer. It cobbles together a good enough solution to a problem from the raw material of solutions to other problems. The solution needs to work, not become Miss Universe.

[2] This sentence really should have been redacted, but it slipped through. Beta DeLoss will make sure that such glitches will no longer occur in future.

The moment we have been waiting for: Enter Alice

Alice: But maybe there is another explanation.

Theo: (Alarmed.) You don't mean ... ?

Alice: (Laughs.) No.
Emily mentioned that biological switching needs to be robust. Maybe the messiness, intermediaries, and the "other variables involved" help in making the system's responses more robust?

Theo: This doesn't necessarily contradict my explanation; both effects may play a role.

Alice: I agree.

Emily: But Alice, how could you support your explanation by evidence?

Bob: This is all getting very philosophical. Could some evidence for Alice's theory be obtained from mathematical theorems?

The others: What would such theorems look like???

(Silence. Everybody thinking deeply.)

Bob's brain on fire

Bob: (Whispers to Emily.) Isn't there more to the biology of a man than gene regulatory networks?

Emily: (Sharply.) Like what?

Bob: Em. ... Thoughts. Feelings. Poetry ...

Emily: Oh, you are talking about the brain.
That's a network of neurons. Each of them can be modeled by diff eq's called the Hodgkin-Huxley equations.

Bob: And these neurons fire!! Right now they fire like mad!!!

Emily: You can think of some region of the state space for each neuron as "firing" and the complement of this region as "resting." This is like assigning a Boolean value to each state of a neuron.

Bob: (Continues.) And this firing is inspired ...

Emily: We say "induced." By the firing of some other, presynaptic neurons. The synapses determine the connectivity of the network.

Boolean brains?

Reka: Could one perhaps model neuronal networks as Boolean networks?

Dave: I have a great class of ODE models for certain neuronal networks, together with a mathematical theorem, like Eve's.

The theorem shows that for each of these ODE models M there exists a Boolean model N that correctly predicts, for a certain subset E of all neurons, their Boolean trajectories, that is, roughly speaking, the order in which they will fire.

Eve: Wow! So the neurons not in E would be, in a sense, intermediaries?

Dave: Exactly so.

Sungwoo: And now we can study how the dynamics of Boolean systems N of that theorem depend on the network connectivity.

Alice wants to know more

Alice: Mathematically speaking, connectivity means, ...

Sungwoo: ... the directed graph (digraph) that you obtain by representing synaptic connections by arcs.

Alice: And for real neuronal networks all these synaptic connections are known???

Sungwoo: Not to any great extent.

Alice: So how can you meaningfully model the connectivity?

Sungwoo: By assuming that these networks are somewhat "typical" and treating them as random digraphs.

Theo: This makes sense!

They were shaped by evolution, which is a stochastic process.

Alice: But can you prove something meaningful about the dynamics under this assumption??

Sungwoo: Yes. By assuming that the connectivity is in a sense completely random, an Erdős-Rényi digraph, together with Dr. Just we were able to obtain a number of interesting results on the transients and attractors in these networks.

Alice: Wow!

Sungwoo: But many interesting open questions remain.

Bob: Listen, pal: There is no open question about my attractor right now and nothing random about my connections ...

Rabi: (Sees that Bob had one drink too many and tries to defuse the situation.) Your brain may be more like a scale-free network.

Bob: Scale-free, that's it!
Oh Emily, a scale from one to ten, nay, from one to one thousand, from one to one million could not even begin to describe ...
(Goes on a tangent and will not be further quoted in this transcript.)

Other types of random connectivities

Rabi: (Chuckles.) Yes, scale-free digraphs do have a few truly exceptional nodes with extremely high degrees.

There is some empirical evidence that connectivities in actual brains are scale-free.

Moreover, brains contain several different types of nodes, and some neuronal tissues, for example, those responsible for organizing visual input, may be structured like small-world networks.

I am currently working with Dr. Just on extending the results mentioned by Sungwoo to networks whose connectivities are multitype Erdős-Rényi digraphs, generic scale-free networks, or small-world networks.

All these types of networks are random digraphs, but drawn from different distributions.

Alice brings us back to the main investigation

Alice: We have two examples, one from gene regulation and one from neuroscience, of ODE models M and corresponding Boolean models N that are consistent in their predictions.

Dr. Y: In my lecture, M would be called a differentiable flow and N would be called a discrete-time dynamical system.
Now we need to precisely define the meaning of consistent.
We did this in our research group several years ago.

Mason: (Happy to explain.) Each variable x_i in the flow M is discretized into two regions that correspond to its Boolean values $s_i = 0$ or $s_i = 1$ in N.
Now at some real times t the Boolean value of some variable x_i switches because it enters the other region.
We choose the times $\tau = 0, 1, \ldots$ in the discrete model N in such a way that they correspond to these switching times in M.
This will assign to each real-valued trajectory in M a Boolean-valued trajectory in N.

Consistency

Alice: But how about consistency?

Mason: In N we have, for each variable s_i a Boolean function f_i like the ones Reka described so that $f_i(\vec{s}(\tau)) = s_i(\tau + 1)$ predicts the next Boolean state of variable s_i based on the current state of all variables in N.
Then M and N are consistent if for a sufficiently large region U of the state space of M, for all trajectories of M that start in U and all such switching times for variable x_i, the Boolean state s_i right after the switch will be correctly predicted by f_i, computed for the Boolean states of all variables right before the switch.

Alice: Is this how consistency works in your model, Dave?

Dave: Not quite, but the idea is roughly the same.
Remember, though, that not all variables in my flow have Boolean counterparts.

Mason: That's fine. We can restrict our attention to only those switches that happen to variables x_i with i in a selected set E.

Strong consistency

Alice: So the variables not in E would then be intermediaries, or, as Reka put it, variables that are merely involved?

Dr. Y: In our group, we called the variables in E signature variables and those not in E signaling variables.

Alice: Eve, is consistency in your theorem what Mason described?

Eve: Yes. But we have something better: At each switching time, **all** of the functions f_i correctly predict the next Boolean state, for **each** variable.
Exactly as in the Boolean networks that Reka described earlier.

Mason: We call this strong consistency. You can get this because your Boolean model is one-stepping, which means that at each time step only one variable changes its Boolean state.
For Boolean models that are not one-stepping, I proved that you cannot get strong consistency with a flow on an open subset U of the state space of M.

Alice wants some toys

Alice: This all sounds very exiting. But do I understand correctly that the flows M in your neuroscience and gene regulatory models are rather difficult to study analytically?

Eve and Dave: Unfortunately, yes.

Alice: So if I want to understand, at a general level, which structural features of a flow M might imply consistency with a Boolean system N and how the signaling variables do or do not help in terms of robustness of the switching of the signature variables, wouldn't it be good to have something like a class of relatively simple toy models for studying these phenomena?

Dr. Y: Definitely. Our group developed and studied two classes of such toy models, one with and one without signaling variables.

Alice: Wonderful!! And what did you find?

Some results

Mason: For the class with signaling variables we proved that we always can get strong consistency with any one-stepping Boolean model, and consistency with any Boolean model that has the weaker property of being monotone stepping.

Ben: And this was confirmed by simulations.

(redacted): Many open questions remain, for example, to what extent "monotone-stepping" is needed in the consistency result.

Alice feels vindicated. Is she?

Alice: And without signaling variables you don't get consistency?

Mason: Nope.

Ben: In simulations without signaling variables, you will see a lot of consistent switching, but also occasional inconsistencies.

Alice: So the signaling variables are needed for robustness! Exactly as I conjectured!!

Theo: Not so fast. Are the interactions between the signaling variables and signature variables in your models designed in a very structured way, or kind of messy as in real biological networks?

Ben and Mason: They follow a fixed pattern.

Theo: So your models don't explain the messiness of biological networks.

How could evolution find such a regular pattern?

(A moment of heavy silence.)

Alice is getting a boost

(redacted): Let's assume for the sake of argument that consistency is a generic property in the sense that it occurs in most networks with certain intrinsic dynamics of the signature variables and somewhat random connections between signaling and signature variables.

Sungwoo and Rabi: Kind of like in the random digraphs that we are studying.

(redacted): If this could be shown, Theo, would you then concede Alice's point?

Theo: Yes. But I doubt whether one could show such a thing.

(redacted): Granted, this may be difficult.

But doing the impossible is kind of fun.

At least I have some ideas of how to go about it.

Alice tries to make a connection

Alice: This sounds so awesome!!!

Can you tell me about your ideas?

(redacted): Will be more than happy to.

Let's move to the free table over there and I will sketch some possible approaches to this research project for you.

Alice: It is getting late.

Let's exchange our cell phone numbers,

(Some totally unexpected technical problem occurred here . . .)

(. . . and the transcript suddenly stops.)

What would happen next remains conjecture.

A memo **From:** Divine Nubes, Committee Chair **To:** The President of Ohio University

This transcript implicitly unmasks several protagonists without authorization.

Legal action against the leaker needs to be taken.

Reka is clearly Réka Albert, Distinguished Professor of Physics and Biology at Pennsylvania State University.

Eve is Eva Gehrmann, now known as Eva Ackermann, author of

Eva Gehrmann and Barbara Drossel (2010); Boolean versus continuous dynamics on simple two-gene modules, *Phys. Rev. E* **82** 046120

Theo is easily recognizable as Theodosius Dobzhansky (1900–1975), a well-known troublemaker.

His unexplained appearance near OU campus raises grave concerns.

A memo **From:** Divine Nubes, Committee Chair **To:** The President of Ohio University

Dave and Sungwoo are easily identifiable as authors of

David Terman, Sungwoo Ahn, Xueying Wang, Winfried Just (2008); Reducing neuronal networks to discrete dynamics. *Physica D* **237** 324–338.

Rabi K.C. is a Ph.D. student at the OU Department of Mathematics.

His full family name remains classified.

Dr. Y, Mason, and Ben are clearly authors of

Winfried Just, Mason Korb, Ben Elbert, and Todd Young (2013); Two classes of ODE models with switch-like behavior, *Physica D* **264** 35–48.

A memo **From:** Divine Nubes, Committee Chair **To:** The President of Ohio University

We are aware of reports by the dishonest media about the identity of (redacted).

However, the transcript clearly shows that on a perfectly fine Saturday night the protagonist spent time at a considerable distance from the office 315C Morton Hall.

This demonstrates beyond reasonable doubt that these reports are fake news.

About Bob, Emily, and Alice we only know that at the end of March they were fired from a shady but apparently well-connected organization, called "neuronal network."

We urgently recommend close surveillance of this organization and strict curbs on its dangerous activities.