Why Is Mathematical Biology So Hard?

Michael C. Reed

Although there is a long history of the applications of mathematics to biology, only recently has mathematical biology become an accepted branch of applied mathematics. Undergraduates are doing research projects and graduate students are writing Ph.D. dissertations in mathematical biology, and departments are trying to hire them. But what should the Ph.D. training consist of? How should departments judge work in mathematical biology? Such policy questions are always important and controversial, but they are particularly difficult here because mathematical biology is very different from the traditional applications of mathematics in physics. I'll begin by discussing the nature of the field itself and then return to the policy questions.

Where's Newton's Law? The phenomena that mathematical biology seeks to understand and predict are very rich and diverse and not derived from a few simple principles. Consider, in comparison, classical mechanics and continuum mechanics. Newton's Law of Motion is not just a central explanatory principle; it also gives an immediate way to write down equations governing the important variables in a real or hypothetical physical situation. Since the Navier-Stokes equations express Newton's Law for fluids, they are fundamental and have embedded in them both the fundamental principle and the complexity of the fluid phenomena that we see. Thus a pure mathematician who proves a theorem about the Navier-Stokes equations and an applied mathematician who develops new numerical tools knows that he or she has really contributed something. Alas, there are no such central fundamental principles in biology. There are principles of course—some would say

dogmas—such as "evolution by natural selection", "no inheritance of acquired characteristics", or "DNA → RNA → proteins". But these are not translatable into mathematical equations or other structures without hosts of additional facts and assumptions that are context-dependent. This means that mathematical biology is very unsatisfying for pure mathematicians, who usually are interested in discovering fundamental and universal structural relationships. It also means that there is no "mathematics of biology" in the same way that ordinary differential equations is the mathematics of classical mechanics and partial differential equations is the mathematics of continuum mechanics.

Diverse, yet special. Because of evolution, biological systems are exceptionally diverse, complex, and special at the same time, and this presents several difficulties to a mathematician. The first is choosing what to work on. There's too much biology! How do changes in the physics or chemistry of a particular environment affect the species that live in the environment (ecology)? How do diseases spread within a population (epidemiology)? How do the organ systems of the human body work (physiology)? How do the neurons in our brain work together to allow us to think and feel and calculate and read (neurobiology)? How does our immune system protect us, and what are the dynamical changes that occur when we are under attack by pathogens (immunology)? How do cells use physics and chemistry to accomplish fundamental tasks (cell biology, biochemistry)? How does the genetic code, inscribed in a cell's DNA, give rise to a cell's biochemical functioning (molecular biology and biochemistry)? How do DNA sequences evolve due to environmental pressures and random events (genomics and genetics)?

The second difficulty is that a priori reasoning is frequently misleading. By "a priori reasoning" I

Michael C. Reed is professor of mathematics at Duke University. His email address is reed@math.duke.edu.

mean thinking how we would design a mechanism to accomplish a particular task. As a simple example, both birds and planes burn something to create energy that can be turned into potential energy, and both have to use the properties of fluids that are implicit in the Navier-Stokes equations. But that doesn't mean that one understands birds if one understands planes. To understand how a bird flies, one has to study the bird. Modelers are sometimes satisfied that they have created a mathematical model that "captures" the biological behavior. But that is not enough. Our purpose is to understand how the *biological* mechanisms give rise to the biological behavior. Since these biological mechanisms have been "designed" by evolution, they are often complicated, subtle, and very special or unusual. To understand them, one must immerse oneself in the messy, complex details of the biology; that is, you must work directly with biologists.

Thirdly, different species (or different tissues or different cells) may accomplish the same task by different mechanisms. An astounding array of special mechanisms allows animals to exploit special niches in their environments. For example, a diverse set of locomotory mechanisms are used at different size scales. Thus, when you have understood bird flight completely, you have not even started on the butterfly or the fruit fly. So, even when one is successful, one may have provided understanding only in particular cases.

We can already draw some conclusions. Don't do mathematical biology to satisfy a desire to find universal structural relationships; you'll be disappointed. Don't waste time developing "methods of mathematical biology"; the problems are too diverse for central methods. What's left is the biology. You should do mathematical biology only if you are deeply interested in the science itself. If you are, there's lots of good news. We mathematicians are experts at thinking through complex relationships and formulating scientific questions as mathematical questions. Some of these mathematical questions are deep and interesting problems in pure mathematics. And most biologists know that the scientific questions are difficult and complicated, so they want our help. There's some bad news too; there are three more reasons why the field is so hard.

The problem of levels. In many biological problems one is trying to understand how the behavior of the system at one level arises from structures and mechanisms at lower levels. How does the coordinated firing of neurons give rise to the graceful motion of an arm? How does the genetic code in DNA create, maintain, and adjust a cell's biochemistry? How does the biochemistry of a cell allow it to receive signals, process them, and send signals to other cells? How does the behavior of groups of cells in the immune system give rise to the overall immune response? How do the properties of individual bees give rise to the behavior of the hive? How do the cells in a leaf "cooperate" to turn the leaf towards the sun? How does the varied behavior of individuals contribute to the spread of epidemics?

We are familiar with these types of questions from physics. What are the right variables to describe the behavior of a gas, and how do the values of these variables arise from the classical mechanics of the molecules making up the gas? The behavior at the higher level is relatively simple, and Newton's law suggests the few important variables at the lower level; even so, the proofs are not easy. In the case of biological systems these questions are even more difficult, because the objects at the lower level have been designed by evolution (or trained by feedback control; see below) to have just the right special properties to give rise to the (often complicated) behavior at the higher level. And it is usually not easy to decide what the important variables are at the lower level. If your model has too few, you will not be studying the "real" biological mechanism. If your model has too many, it may be so complicated that a lifetime of computer simulations will not give new biological understanding. You need ideas, guesswork, experience, and luck. You need to be able to deduce the consequences from the assumptions. That is what mathematicians are good at.

The difficulty of experimentation. We mathematicians often have an overly simple view of experiments and the role they play, probably because we don't conduct them ourselves. A theory is tested by deciding on a few crucial variables and designing the right experimental setup. For example, one measures how fast metal beads and feathers fall in a vacuum or the angle subtended by two stars. However, the complicated histories of interaction between theory and experiment in quantum mechanics, nuclear physics, and elementary particle physics in the twentieth century show that this simple view is naive. And for several reasons the experimental situation is even more difficult in biology.

First, one is often interested in how the behavior at one level arises from lower levels. Typically, this "emergent" behavior cannot be seen in any of the parts at the lower level but arises because of complex interactions among the parts. Unfortunately, it can be misleading to study the parts in isolation. For example, I try to understand how certain biochemical networks in mammalian cells function. The networks give rise to systems of ODEs in which the nonlinear terms depend on the enzyme kinetics for each separate enzyme. The enzymes can be isolated and their reaction kinetics studied "in vitro" in experiments that combine

pure enzyme with pure substrate. But in the soup, in the real cell, the enzymes and substrates are binding to or being affected by other chemicals too, so one is unsure whether the "in vitro" experiments reflect the true "in vivo" kinetics. That is, each of the parts at the lower level behaves differently in isolation than when it is connected to the other parts.

Second, chance plays a role, not only in experimentation, but perhaps in explanation. Why are two neighboring fields dominated by black-eyed Susans and poppies respectively? Are the soils different? Are the local plant and animal species different? Or perhaps the "explanation" is a chance event in the past (or all of the above).

Third, individuals (whether cells or flowers or people) are both similar and different. How does one know whether data collected is "special" or "typical"? How does one assure oneself that rat data tells us something about humans or, indeed, something about other rats?

Finally, it is characteristic of living systems that the parts themselves are not fixed but ever changing, sometimes even affected by the behavior of the whole (see feedback control, below). A simple true story illustrates this point. An experimenter (#1) who used rats in his experiments was getting very unusual results, and the results were not repeatable week to week. After six months of this he had to stop his experiments and investigate the rats. His rats were housed in his university's vivarium. It turned out that another experimenter (#2) had a mean technician, and when the other experimenter's technician came to get #2's rats, #2's rats would cry out, upsetting the rats belonging to #1. The behavior of #1's rats in experiments depended on whether #2 was doing an experiment the same day! Whew, I'm glad I became a mathematician.

For all these reasons, biological data must be approached cautiously and critically. Since biological systems are so diverse and everything seems to interact with everything else, there are many possible measurements, and enormous amounts of data can be produced. But data itself is not understanding. Understanding requires a conceptual framework (that is, a theory) that identifies the fundamental variables and their causative influences on each other. In messy biological problems, without simple fundamental principles like Newton's Law, useful conceptual frameworks are not easy to propose or validate. One must have ideas about how the structure of (or behavior of) the whole is related to the assumptions about the parts. Thinking through such ideas and proving the consequences of the assumptions are important ways that we mathematicians can make contributions.

The problem of feedback control. It is common to think of biological systems as fragile.

However, most are very stable, and it is almost a tautology to say so, because they must all operate in the face of changing and fluctuating environmental parameters; so if they weren't stable, they wouldn't be here. We are familiar from engineering with the concept of feedback control, whereby variables are sensed and parameters are then reset to change the behavior of the system. The nephrons in the kidney sense NaCl concentration in the blood and adjust filtration rate to regulate salt and water balance in the body. The baroreceptor loop regulates blood pressure, heart rate, and peripheral resistance to adjust the circulation to different challenges. Numerous such control systems are known and studied in animal and plant physiology.

There is another kind of feedback that operates between levels that poses special problems. Here are two examples. In the auditory system sensory information is transformed in the cochlea to electrical information that proceeds up the VIIIth nerve to the cochlear nucleus and from there to various other nuclei in the brain stem (a nucleus is a large anatomically distinct group of cells) and on to the midbrain and the cortex. Surprisingly, there are also neural projections from the cortex that influence the sensitivity of the cochlea. Second, the dogma DNA → RNA → proteins → function has turned out to be a naive fiction. Genes (pieces of DNA) don't turn themselves on or off but are activated or inhibited by proteins. That is, proteins affect the genes, adding a reverse loop to the simple picture, implying of course that the genes affect each other through the proteins. The "fundamental" objects, that is, the objects most closely related to "function", may not be genes or proteins but small networks involving both genes and proteins that respond in certain ways to changes in the cell's environment. These kinds of examples show that the nineteenth-century picture of a machine with parts is a very inappropriate metaphor for (at least some) biological systems. When there is feedback between levels, it is hard to say which are the parts! In fact, it may be hard to say which are the levels, and therefore our traditional scientific research paradigm of breaking things into smaller parts (lower levels) may not be successful.

This is not just a philosophical point but a fundamental research issue that deepens the impact of the previous four difficulties. Take, for example, the question of dendritic geometry. It's been one hundred years since Ramon y Cajal made beautiful drawings of complicated dendritic arbors on nerve cells. Is the geometry important? Surely it must be, we feel, since cells in the same brain nucleus in different individuals seem to have roughly similar dendritic arborization. And, indeed, there are examples where it is understood how specific dendritic geometry creates specific neuron-firing properties and presumably specific cell function,

though it is not always clear what "function" means for a single cell embedded in layers and layers of a large neural network. On the other hand, suppose a cell is part of a large neural network whose job it is to transform its pattern of inputs into a corresponding pattern of outputs. This neural network may have been trained to do this job by feedback control from a higher level, in which case the details of the dendritic geometry (and even the details of the neural connections) may not be important at all. The details arose from the training, and they are whatever they need to be to give the behavior at the higher level. Furthermore, for large networks there may be many choices of details that give the same network behavior, in which case it will be hard to infer the behavior of the whole by studying the properties of the parts.

I now want to turn to the policy questions that I mentioned at the beginning. I have been using the term "mathematical biology" to refer in the broadest way to quantitative methods in the biological and medical sciences. Physicists, chemists, computer scientists, and biological and medical researchers with some mathematical training can and do contribute to the field I have been referring to as "mathematical biology". But let us now narrow the focus to mathematics education, both undergraduate and graduate, and the mathematics job market.

Undergraduate education. Mathematical biology is an extremely appealing subject to undergraduate students with good training in freshman and sophomore mathematics. Many are naturally interested in biology, and all know that we are in the midst of a revolution in the biological sciences. They are usually amazed and delighted that the mathematical techniques that they have learned can be used to help understand how biological systems work. Further, mathematical biology is a perfect subject for undergraduate research projects. Biology is so diverse and so little quantitative modeling has been done that it is relatively easy to find projects that use undergraduate mathematics in new biological applications. The students find such projects to be very rewarding. They know that the undergraduate major consists mostly of nineteenth-century mathematics; of course they are excited by twenty-first-century applications. Here at Duke we have found that the availability of projects in mathematical biology has attracted many students to the mathematics major. Of course, it helps to have a mathematical biologist on the faculty, but it is not necessary. Any mathematician can create and supervise such projects by working cooperatively with local biologists. It requires only the effort to make the connections and tolerance for appearing "nonexpert" to the students (something we are not used to!).

Graduate education. There is quite a bit of disagreement about the proper mathematics graduate training of a student who wants to be a mathematical biologist. I'll simplify the discussion into two (extreme) positions. The first position emphasizes maximal contact with biology and biologists as part of graduate training. Graduate students should take biology courses (including labs) and should participate in or even initiate collaborative modeling projects. This way they learn a lot of biology, and, even more importantly, they learn modeling and how to communicate with biologists. By doing this they study less mathematics of course, but they can learn what they need later when they need it. The second position emphasizes training as a mathematician first. Graduate students should receive the traditional training in analysis or applied mathematics (or other subjects), and (ideally) the thesis should contain some applications to biological problems. But the graduate student should not spend too much time slogging around in the biological details or working on collaborative projects. It is the job of the thesis advisor to be the interface between the graduate student and the biology. Later, after the Ph.D., when the mathematician is established, he or she can choose to become involved in collaborations and learn more biology.

I guess that most mathematical biologists support the first position. I support the second, perhaps because that is the route that I followed myself. Mathematical biology is really a very hard subject (I hope I have convinced you of that), and a great many ideas and techniques from different branches of mathematics have proven useful. So mathematical biologists need broad training in mathematics. Secondly, I believe that only deep and rigorous graduate training creates mathematicians who can not only learn new mathematics when they need it, but who can also recognize what they need to learn.

Hiring issues. Hiring a mathematical biologist poses special challenges for departments. Most mathematicians have no idea how large the field "biology" is or how large the research communities are. Two examples illustrate this. Here at Duke, Arts and Sciences has 469 tenure-track faculty members, and the Medical Center has 767 (not counting clinical faculty). My colleague Harold Layton is a mathematician who works on the kidney, so of course he goes to the annual meeting of the American Society for Nephrology, where the typical registration number, 12,000, completely dwarfs the registration at the Joint Mathematics Meetings. And those constitute just a subset of the researchers who work on the kidney! So, the first issue is thinking about what kind of mathematical biologist you want. It is a good idea to involve local biologists and medical researchers in preliminary discussions, both to educate department faculty and to understand the local context.

The question of how best to judge job candidates is particularly difficult in mathematical biology. First, most mathematicians know less biology than their tenth-grade daughters. That's just the way it is; biology was not a mathematically related discipline when we were growing up. But, more importantly, mathematical biology really isn't a "field" of mathematics with a coherent community that can testify meaningfully about young people. Mathematical biology is fragmented because biology itself is so diverse. A mathematician working on the lung might want to talk to pulmonary physiologists or geometric analysts who are experts on fractals, but why would he or she want to talk to mathematical biologists working on the kidney, the neurobiology of hearing, or the epidemiology of AIDS? In each area of specialization there is a tremendous amount of biology to learn, and if one doesn't have the background, it's hard to judge the strength of an individual's contributions. This is true even for us, the mathematical biologists, the "experts" you expect to consult. So hiring in mathematical biology necessarily involves intuition and high risks as well as high potential payoffs for the department and the college.

The first step. The best way for departments to overcome these difficulties is to encourage senior faculty to become involved in bringing biological applications and student projects into the undergraduate curriculum. This should be done by working cooperatively with local biologists to create examples and projects related to their own specialties. All mathematicians can do this, and they do not have to give up their own research agendas or become mathematical biologists; it only requires effort.

This strategy for engagement with biology has great benefits both for departments and individuals. The faculty as a whole will become educated in biology and thus better able to judge job candidates in mathematical biology, and the undergraduate curriculum will be more attractive. More importantly, departments and individuals will be participating intellectually in the biological revolution, the greatest scientific revolution of our times, perhaps of all times. The task is to understand how life, in all its diversity and detail, works. This includes how we act, think, and feel, and how we influence and are influenced by other forms of life. We mathematicians have the technical and intellectual tools to make enormous contributions. So, surely, this is our responsibility and our opportunity.