PROFESSOR: So, welcome to systems biology. There are many different numbers that you might have signed up for this class through. My name is Jeff Gore. I'm an assistant professor in the physics department. And I think that this is really-- it's a fun class to teach. I hope that those of you that stick around for the rest of the semester find that it's a fun class to take as well.

I just want to give a little bit of an introduction of the teaching staff, and then we'll go over some administrative details before-- most of today, what we'll do is we'll basically just spend an hour. I'll give you a flash summary of the course, and then you'll have a sense of the kinds of ideas we'll be exploring.

So, what is this thing, systems biology? So it's, I would say, ill-defined. But in general, we have this idea that, in many cases, the really exciting functions that we see in biology are arising from the interactions, at a lower level, of relatively simple components. So, the kinds of behavior that we'd like to understand are nicely encapsulated in this video of a neutrophil chasing a bacterium. And if I can-- if my computer is actually doing something, then-- there's something.

So, this is a classic video that many of you guys might have seen over the years. So, this was taken in the 1950s. It's, as I said, a neutrophil, which is part of your innate immune system. So, sort of the first line of defense. When you get a bacterial infection, for example, this white blood cell is going to chase the bacterium and eat it up before it can divide and harm you.

So, I'm going to play that over again, because it's pretty cool. So, the features that you want to try to pay attention to-- so, this is a single cell that somehow is able to track this bacterial invader. It's using chemical cues in order to figure out where it is

that bacterial cell is going. It's able to disregard these red blood cells that are in its way, push them aside, change direction, before it eventually captures this bacterial cell. So, all of these are striking, amazing information-processing capabilities, where that information has to be coupled not only to some sort of decision-making within that cell, but also, it has to be transduced into these mechanical forces and motions that allow that cell to capture the bacterial cell. So, let's try it again.

So, here it is. You can see the bacterial cell is here. So, it's ignoring this other cell, keeping focus on this one, pushing aside the red blood cells, so it can follow it along. Every now and then-- now the cell changed direction. But eventually, you can see it catches up to the bacterial cell and eats it. Now you're not going to get sick from that infection.

So, that's an example of the kinds of remarkable behavior that can be implemented even just by a single cell. So, we know that, as humans, we have brains with 10 to the 12 neurons or so. So, maybe you'd say, it's not surprising that we can do fancy things. What's remarkable is that even at the level of an individual cell, it's possible to implement rather sophisticated information processing capabilities. So, this is the kind of thing that we'd like to be able to say something about by the end of the class.

Many of you, I think, probably read the course description, and this gives you a sense of the kinds of topics that we're going to be covering over the course of the semester. I'm not going to read it to you. But one thing I want to stress is I brought up this general idea of systems biology. How is it that function arises from interactions of smaller, simpler parts?

I think it's very important, right at the beginning, to be clear that there are really, I'd say, two distinct communities that self-identify as studying systems biology. And they're-- to simplify it a bit, what I would say is that, basically, there's the physics, or physics-inspired community, where-- and I'm in the physics department, so that's where I fall, and where this class is going to be. So, it's really trying to use some simple models from nonlinear dynamics or stochastic processes combined with

quantitative experiments, often on single cells, in order to try to illuminate how this cellular decision-making process works. And on over the next hour, you'll see-- get a flavor of what I mean by this.

Now, there's another branch of systems biology that is also very exciting, and that many of you may want to learn more about in the future. And maybe some of even were thinking that this was what the class is going to be in. But let me explain what is.

This other branch, I'd say, is more influenced by computer scientists and engineers, where what they're really trying to do is use complex models, machine learning techniques, and so forth, in order to extract signal from large data sets. And this is also, again, systems biology because it is trying to understand how the global properties of the cell result from all these interactions, but it's a rather different aesthetic take on the subject. And, indeed, much of the activity, that one would be doing here is different from the more physics branch of systems biology.

So, if what you were looking for was more of this large data set, high throughput, branch of systems biology, then you may not be at the right place. And there's an interesting fact, which is that if you decide that you want this other branch of systems biology, then, very conveniently, you actually have space in your schedule, because Manolis Kellis is teaching a class in computational biology-- really, this other branch of systems biology-- at the same time. So, if you think you're at the wrong place, you're welcome to just sneak out now. Go to over 32-141, and I'm sure that he will welcome you, and no hard feelings.

Similarly, in the spring, there's another computational biology class that some of you may be thinking about. And this is taught by Chris Burge and company, and also, I'd say, maybe more of this other branch of systems biology. And, finally, once again, in the spring there's a class, quantitative biology for graduate students, that I'd say probably assumes somewhat less mathematical background. So, if after looking at the syllabus, or maybe even getting started looking at the first problem set, if you think that maybe this class is expecting too much, then you may want to consider

taking quantitative biology in the spring, and maybe taking systems biology-- this class-- next fall.

So, on that note, I want to say something about the prerequisites. The major challenge with this class-- certainly, teaching it, from my standpoint, and I think for many of you taking it-- is that there's a wide range of different backgrounds. So, we can maybe get a sense of that now. Just a show of hands, how many of you are undergraduates? So, we've got a solid third or so. How many are-- mixing together undergraduate and graduate-- but how many of you are in the physics department at one level or another? So, we have maybe, again, a third.

Biology department? So, we have a quarter, a fifth. Engineers? So, a substantial fraction of engineers. And chemists? We've got a few of them. Mathematicians? All right, we've got one.

So, if you did not raise your hand, where you based, physically, intellectually, something? Did we get everybody? OK.

So, you can see that there's a really broad range of different backgrounds. And what that means in a concrete way for us is that I will very much try to avoid using unnecessary jargon or unnecessary mathematics. I think that mathematics is a wonderful thing, but in some cases it, I think, obscures as much as it illuminates.

So, for me, I very much try to focus on conceptual understanding. And on top of that, I try to build-- you like math, too. But I think that it's very important to be able to, for example, plot your solution. So, after you derive some fancy equation describing something, you should know whether that thing goes up or down as a function of something or another. And I think it's very easy to lose sight of these basic aspects when we get too deep into the mathematical equations.

But, that being said, we do, I'd say, expect something-- not necessarily full-- you don't have to have taken the full class, 701 or 702, but at least a solid high school class of biology. If it's been more than 10 years since you took a biology class, you might want to take one before coming here. You could, in principle, catch up, like all

things.

We also assume some comfort with differential equations and probability. So, we've actually added those as prerequisites, particularly from the standpoint of an undergraduate to give you a sense of the sort of material that we expect you to be comfortable with. So, we will not be defining probably distributions so much. We will assume that you can calculate means and standard deviations of discrete, continuous distributions, and so forth.

And the other thing that is going to be important is that a major goal of the class is to increase your comfort level with using computational techniques to analyze some of these problems. So, every week, we're going to have a problem set. And on every problem set, there will be at least one problem where you have to use some computational package in order to calculate something. So, are you going to do a simulation to understand the stochastic dynamics of this or that? Or maybe you're going to integrate some differential equations.

And in this case, you can use whatever package you like. So if you are a MATLAB person, that's fine. Mathematica is fine. The officially supported language is going to be Python, because that's what Sarab-- if he's going to be spending hours helping you with your code, he wants it to be something that he's comfortable with. So, that's going to be what we might call the official language, in the sense that he will perhaps provide some sample code and so forth to get you started. But you're welcome to use anything that you want.

And that being said, we will have a Python tutorial almost certainly on Monday. We're waiting to get-- to find out what the classroom is going to be. But we will send out a notice to the class about that, as well as instructions on how to get Python on your computer.

Are there any questions about where we are so far? What we've said? Expectations, prerequisites? All right.

So, grading. One of the things that we have to do is we have to grade. I'd say that

our goal is very much to help you learn material that we're excited about. So, I am not in any way trying to grade in any mean way. And what that means is that-- we also just don't want you guys to feel like you're competing against each other.

So, what that means is that-- so, these are the grade cutoffs. So, they will not-- numbers will not go up from here. If I screw up and I make some really hard exam, then I reserve the right to lower these numbers. But basically, this is what's worked for the last several years. So, you should feel comfortable collaborating with your friends to study, to try to figure out the material, because your grade is just going to be determined by where things end up on this chart, basically.

And the course grade is going to be split, as you can see. There's a fair component on problem sets, and that's because the problem sets are-- they're going to be hard work. We're going to have problem sets every week, and you can expect to spend a significant amount of time on them. And the thing that you learn in doing these computational problems is somewhat different from what you learn and what you demonstrate on an exam. So, that's why it's not all just an exam grade.

I'm going to say something more about these pre-class reading questions. There are going to be two midterms and an exam. The dates are on your syllabus, so please mark these evenings on your calendar.

So, the problem sets. You can read about this. But basically, every Friday at 7 o'clock, they're going to be due. A box out between the third floor of building six and the fourth floor of building 16, I suppose-- these are the physics homework boxes.

So, the idea is we'd like you to have a weekend to catch up and start reading for the next week. So, that's why they're dude just before dinner on Friday. That being said, we understand that sometimes there are a lot of problem sets, or sometimes you're overwhelmed with something else. So, that's fine. You can turn it in for 80% credit till Monday morning at 10:00 AM, when we're going to post the solutions. So, we won't be accepting problem sets after that, unless you get agreement from Sarab in advance.

So, the pre-class reading questions. I'd say this is a key part of the class. It's only 5% of the grade, and it's graded really only on participation-- that you've done it. But this is an essential element of what we like to call a flipped classroom.

So, today's class is going to be rather different from the rest of the semester in that today is more like a lecture, I would say. Whereas the rest of the semester, there will not be any PowerPoint slides and it'll be very much, I hope, very interactive. In order to facilitate that, there are a number of different elements. One is that we do require reading before class. And the way that we encourage you to do the reading is that we ask you to answer questions the night before.

So, what you're going to do is you'll, by 10:00 PM, the night before, just three questions-- just a couple of sentences each question. It's not that you're supposed to have to do a lot of work. It's just that if you did the reading, you should be able to give your take on it, and you think about it a little bit. And then Andrew will go over the submitted answers, and we'll send out his favorite answers among the group. So, your answer will occasionally be represented there, if you say something that's reasonable.

Now, it's really important to have done this reading and thought about the material some before, because the idea is that, in class, we'd really like to engage in what you might call some higher level learning. So, it's not just this idea that-- the traditional lecture arose when books were very expensive. So, if you're at a university in the 13th century, you don't have a textbook. So, what you need is for me to stand up front and read to you.

And that's fine, except that it's better for you to read it. And you can read it outside of class, think about it a bit. And then that means when you come to class, we can actually discuss it. In particular, I'll give you my take on the material, the research that I'm excited about in the area that's been published recently.

And we will also try to get you involved via these concept questions. So, in future sessions we'll have these flash cards, or these colored cards, so we can ask these conceptual questions. A, B, C, D-- if you drop an apple, does it go up, down, left,

right? And then, you guys get to vote. And then, after the vote, we will often have you pair up.

And the goal there is that you're trying to convince your neighbor that you're right. And after that, you might expand it to fours or so. But the idea there is that it's very important for you to try to confront the material, make your best guess, and then discuss it with a neighbor. And I think that this is actually one of the fun aspects of the course. At least, I think so.

And I'll say also that this basic technique is the result of-- there's a whole field of education research. And there are very, very consistent and strong signals in this, suggesting that this sort of flipped classroom, active learning style, actually is good for learning. So I'm not just doing this because-- it is more fun, but that's not actually why I'm doing it. I'm doing it because the people who have spent their lives studying this topic have included this is the best way to teach.

Any questions about the pre-class questions or my notion of active learning? All right.

You can mark your calendars in advance. We do have a final. It has not been scheduled yet. It'll be sometime the week of December 15-19. So, for those of you who are looking online for plane tickets back home, after December 19. Or, if you'd like, you can wait a couple weeks, and then the final will be scheduled.

We have two required textbooks for the class. The first is *An Introduction to Systems Biology* by Uri Alon. I think it's a wonderfully clear, exciting introduction to the topic. The flip side of being wonderfully clear is that it's a little-- you could complain that it's too simple. And what that means is that we will be supplementing the book in a variety of ways, both with separate notes, and also by extensive reading of papers from the primary literature.

The second half or third of the class, we'll be reading some chapters from *Evolutionary Dynamics*, a book by Martin Nowak. Again, a very nice, I think, clear, exciting introduction to that field. So, I think that these are both books that, if you're

at all interested in this area, you should own anyways.

There are two other books that you might want to recommend buying. So, first there's *Essential Cell Biology*, which is kind of like the easy version of the cell, also by Alberts. So, be careful of just buying a book by Alberts. So, I'd say *The Cell* is everything you ever wanted to know about the cell-- and more than you want to know about the cell-- whereas *Essential Cell Biology* is really just a wonderful book.

We read this in my lab as kind of a summer book reading project, where each week, we read a chapter, and we got get over lunch. And we just went around the table, and we went through all the questions in the book-- really. And we just alternated, and we discussed, and it was-- really, it's wonderful. It focuses on the ideas. You have to memorize a few things. But every now and then, you need to memorize something in order to keep track of what's going on. But I would say that if you're really interested in biology in any serious way, then I would recommend you buy this book.

And then, finally, there's this book *Nonlinear Dynamics and Chaos* by Steven Strogatz, which is a beautiful introduction to denominator dynamics. If you have not seen the book, I encourage you to check it out. And in particular, some of the topics on stability analysis, and oscillations, bifurcations, and so forth-- this is a really great way to learn about them.

I want to just give a brief plug for-- there's another class that some of you, especially the first year students interested in biophysics, might be interested in. This is 8.590J slash 20.416J slash seven something. So it's a class targeted for first year graduate students interested in biophysics.

Basically, each week, we read a paper. We have a different guest lecturer come from across campus, either physics, chemistry, biology, biological engineering, civil or environmental engineering. So, a great way to meet different faculty who are working in the interface of physics and biology in one manifestation or another. The class-- it's going to be this Friday from 3:00 to 5:00 PM. But then, in later weeks, it's going to be 4:00 to 6:00 PM because we-- because it conflicted with something.

So, I'm going to tell you-- I'm going to give you the overview of the rest of the semester in terms of the science. But I just want to first remind all of you that, starting on Tuesday, it's going to be the real class. What that means in particular is that we expect you to have done some reading, and we expect you to have submitted your pre-class reading questions by Monday night at 10:00 PM. We used to have it at midnight, but then Andrew has to stay up really late to go over all your responses and send out them-- so, 10 o'clock. And then we'll get going on simple interactions between doing enzyme and substrate, simple gene expression ideas, and so forth.

So, I'd say that the course has three parts. There's like-- the first half is part one, and then part two is the half to 3/4 mark, and then the last part is maybe four or five lectures. And the structure of this is really-- it is going from the microscopic scale, and then-- in terms of just the basic ideas of, what happens if molecule A binds with molecule B. What are the features that we should be aware of? So, pretty basic there.

All the way up to questions in ecology. The last lecture is going to be questions about the origin of diversity in ecosystems. So, we'll basically march from the molecular scale up to the population scale throughout the semester.

For those of you who are interested in thinking about these questions of how to organize a class, there's quite an interesting discussion at the beginning of Bill Bialek's *Biological Physics* book, where he very explicitly says that he tried to resist the temptation to do what it is that we do in our class. He resisted the temptation to start from the small scale and then build up to these larger scales. And the reason he says he wants to avoid that is because he does not want to give students the impression we actually understand how you go from the lower scales up to the higher scales. And I think that's a totally reasonable viewpoint.

But that being said, the whole point of this endeavour is to try to say something about it. We may not really understand it all, but we have to try. And it's certainly true that is how function arises, that there are lower level interactions that lead to

higher scale functions, dynamics, behaviors. We may not be able to predict exactly what's going on there, but that is the way that nature does it. So, I don't want a second guess nature, certainly.

So, that's going to be our approach. But if at the end of the class, you prefer a different order, you can always just turn yourself around, and then it all jumble up in your brain. And then, it can be whatever order you like.

On Tuesday, we're really going to start with the most basic ideas. What happens if you have, for example, one gene that is going to turn on another gene? So, you might have a transcription factor X that's going to activate so gene Y, that says it's going to cause gene Y to be expressed. Now, it's as simple as you can get. But what are the general features that you can say about this sort of process?

Well, you can say, there's one thing. It's that X could either be an activator of Y, or maybe it's a repressor of Y. So, these are the two symbols that we'll often use. An arrow will either be an-- well, this symbol will always be a repressor. A plain arrow may be ambiguous, so beware.

Now, the question is, what happens here? For example, you might have this transcription factor-- let's say 10R-- that's repressing expression of this gene that is encoding GFP, green fluorescent protein. You're going to see this many, many times over the course of the semester. In some ways, one of the things that we're going to see in the class is that new ideas often arise from new techniques or new capabilities.

Now, it was really, I think, the Y-- the spread of GFP and related proteins that allowed us to visualize gene expression in individual cells. And it led to this real flowering of new ideas, of how, for example, stochasticity may be relevant, cell to cell heterogeneity. These are all, I think, very interesting ideas. But in order for them to be concrete, you need data. And this was a powerful way for us to get data that was relevant for these sorts of big questions.

So, the idea here is that if this protein is made, it's expressed. Then that cell will

become fluorescent. In particular, it'll become green if you shine the proper light on it. And then, we can do things. We can ask questions about, for example, the dynamics of this process.

So, here we have a case where, now, this is a repressor that is-- if you have this repressor, then it stops expression of that fluorescent protein. Now, you can ask, what happens if you start in a situation where the cell is repressing expression of that gene. So, in this case, the protein concentration is 0, so the cell is not fluorescent. But then, you add something so that now you cause that repressor to fall off and stop repressing that gene.

Now, the question is, how long does it take for the protein concentration to grow to some equilibrium? It starts out at 0. Eventually, it's going to reach some steady state. So, what is it that sets this time scale? What's the characteristic time that it takes for the cell to respond to this signal?

What we're going to find is that there's a very general sense in which that characteristic time scale is really the cell generation time. So, cells divide at some rate. It depends on the kind of cell, the environment, and so forth. Does anybody have a sense for a bacterial cell in nice, rich media, good temperature-- how long it takes for it to divide?

Yeah, 20 minutes. So, E. coli, for example, can divide every 20 minutes, if you put it in the right environment. Which is really an amazing thing, if you think about the number of different proteins that have to be made, and the complicated mechanics of growing, and separating, and so forth. But every 20 minutes such a cell can divide.

That's saying that a bacterial cell, when it sees a new signal, it's going to take, of order, that amount of time in order for it to do anything. And that's just because of this natural process of dilution. So, as the cell grows, there's a dilution of the contents. So, it makes sense. If you start out with a protein and you stop making it, then maybe you'll get an exponential decay of that concentration with this time scale, the cell generation time.

12

What's interesting is that, in some ways, that resolve is more general. That, even if you're trying to turn something on, there's the same limit, this cell generation time, that is placing some limit to how fast the cell can respond to new information, if it uses this mode of information transmission where you express a new gene. So, if you want to go faster, than you have to do something else.

So, in some cases, you can actually have a situation where a protein is actually regulating itself. So, this is an example of what you might call negative autoregulation. So, in this case, that protein actually comes back, and it represses its own expression.

It's found that this is actually rather common in biology. And so, of course, if you see something that is common in biology, then it's reasonable that-- so, maybe there's an evolutionary explanation. Not always, but it gives you a hint that maybe it's worth looking.

Now, in this case, what we're going to find is that such negative autoregulation does some very interesting things. So, for example, one thing that it's been shown to do is to increase the rate of response of that gene. So, in some ways, you can speed up a response to some signal by having that negative autoregulation. In a similar way, this negative autoregulation increases what you might call robustness, the ability of the function-- in this case, maybe, the concentration-- of the protein to be robust to variations in things like the temperature, or this or that. So, environmental perturbations, or maybe just stochastic fluctuations.

Now, in this field, I'd say one of the key advances that led to the birth of both this branch of systems biology, but also the field of what you might call synthetic biology-- really using engineering principles to try and design new gene circuits-- was a pair of important papers that we're going to be talking about in this class. So, the first of these was a paper from Jim Collins's group. He was at BU, although you may not have heard yet, but he's actually just agreed to move over here at MIT. So, this is very exciting for us, and hopefully for you.

So, Jim Collins-- in 2000 he showed that he could engineer a switch, something called a toggle switch. So, if you have two genes that are mutually repressing each other, then this is a system that the most basic memory module. Because if you have one gene that's high, it can repress the other one, and that's a stable state. But if this other gene goes high, then it's going to repress this one here, and that's another stable state. And that state, since it's stable, can maintain memory of the past environment.

And he was able to demonstrate to his group that he could construct such a switch using components that, in the past, were never interacting with each other. So, this is taking advantage of this fabulous modularity of the components of biology in order to do something that, is in, principle useful. And by doing this, it's possible that you could go and engineer new things. But it's also a test bed for you to take this dictum from Feynman that if you can't build it, then you don't understand it.

And this is a nice way to go into the cell and say, if it's really true, if all these models that we talk about in systems biology, for example-- if they're really true, then we should actually be able to go into the cell, put these components together, and demonstrate that there is, for example, this switch-like behavior. And this was a very important paper that demonstrated that it's possible to do this. The other paper that I think--

AUDIENCE: Did they actually make the switch?

PROFESSOR: Yes They actually constructed it. They put it on a round, circular piece like this plasmid, put it into E. coli, and showed that they could do this here. And indeed, this particular issue of nature, I think, was hugely influential for our field, because that toggle switch paper and this other paper-- "The Repressilator," by Michael Elowitz and colleagues-- they were kind of back to back in that issue of nature. And I'd say, in some ways, they were the beginning of systems and synthetic biology. Of course, you can argue about this. But certainly, I think they influenced many, many people in getting excited about the field.

So, the repressilator-- this is the idea that you can generate a gene circuit like this

that will oscillate. And in this case, instead of having just two genes that are repressing each other, if instead you have three genes that are repressing each other, but in a circular fashion, then there's no stable state akin to what we have with this toggle switch. But instead, what happens is that you get successive waves of each of these components going up and down. So, they oscillate as they mutually repress each other. And I just want to be clear about what this is.

So, here, these are E. coli cells, where Elowitz put in this plasmid-- this circular piece of DNA-- encoding those three genes that mutually repress each other. And basically, associated with one of those genes, he's again attached one of these fluorescent proteins. So, the level of fluorescence in the cell tells you about the state of that gene circuit. Let's see if we can--

So, it starts out. There's a single cell you can't really see. It starts dividing. Then you see it oscillates-- gets bright, dim, bright, dim. But you can see that there are a number of features you might notice about this movie. So, first, it does oscillate, which was huge in the sense that it wasn't obvious that you could actually just put these genes together and generate something that oscillates at all.

On the other hand, you'd say, well it's not such a good oscillator. In particular, for example, this started out as a single cell. Now it's dividing under the microscope on agger. So, it's getting some nutrients there. But what you see is that-- are these cells all in phase with each other? No. So, there's patches-- bright, dark.

So, the question is, what's going on here? And it turns out that this design of an oscillator is perhaps not a very good one. And, indeed, one of things we'll be talking about is how you can maybe use some engineering principles to design better oscillators. So, for example, Jeff Hasty at San Diego has done really beautiful work showing that you can make robust, tunable oscillators in cells like this.

Now, these oscillations they were maybe not as good as you would like. But this, actually, is an example of how a partial failure-- in the sense that they're not great oscillations, that maybe somehow there's noise that's entering in here that you would not like. This led to the realization that maybe noise is relevant in decision

making within cells.

And this led-- I'll show you in a few slides-- to another major advance that Elowitz had. So, this is, I think, a good example of how one-- we might call it a partial failure. Some reservations about the quality of this oscillator led him to another really big scientific discovery on the importance of noise in decision making within cells.

But before we get to this noise, we're going to say something about the more global structure of these gene networks. And, in particular, we're going to analyze, and we're going to read this paper by Barabasi which represents a simple mechanism for how you might what are called these power-law distributions in networks. So, if you have these genes that are mutually activating-repressing, what can you say about the structure of this gene network within the cell?

Now, you can analyze such global structure in a couple of different ways. One is just to ask, how many different genes are different genes connected to? And that's maybe the more Barabasi approach.

But then there was another major discovery that Uri Alon, the author of our textbook, made, which is that you can ask-- in this crazy network that you have that describes the decision making within the cell-- are there common patterns or motifs that appear over and over again? So, just like this idea of autoregulation, when a gene represses or activates itself, that's something that appears frequently. So, you can ask, why might that be?

Similarly, if there are other patterns that appear in these networks, then maybe they arose, or they were selected for by evolution, because they perform some other function. In particular, we're going to analyze this feed-forward loop motif, where you have some gene that activates, for example, another gene Y. Y activates-- I'm sorry, this is supposed to be Z. Now, if X, again, directly activates or represses Z-- this bottom gene-- then what does that do for you? Because this is something you see more frequently than you expect, based on some notion of chance, or some null model.

So, the question is, why would these feed-forward loops appear over and over again? And it turns out that they can provide some nice functions in the sense that, for example, you can provide some asymmetrical response to temporary fluctuations of inputs, et cetera, et cetera. So, we'll try to get some sense of these ideas.

So, as I alluded to before in this idea of the repressilator that Michael Elowitz made, he saw that it was surprisingly noisy. And this got him thinking about the role of stochastic fluctuations within cells. And I think that this is a common theme throughout much of systems biology. It's the role of noise in biology.

And this could be within a cell for individual decision making. It could be in context of development-- how is that you robustly make a body, given noise? It could be at the level of evolutionary ecosystems, that maybe noise actually plays a dominant role in, for example, determining the abundance or diversity of ecosystems. So, we'll see these themes pop up on multiple scales throughout the semester.

But in this case, what Elowitz did-- this is just two years after his repressilator paper-- he showed that if you just take the-- in a single cell, you give it the exact same instructions. So, you say, make a red fluorescent protein, and make a green fluorescent protein, with the exact same instructions to the cell. And you can say, well, if you have the same instructions, then the level of the red and the green should do the same thing. But what he found was that, actually, there was surprising heterogeneity of the level of those two proteins, even in single cells.

So, the idea is that even-- this represents a fundamental limit to what a cell can do, because this is saying that we take-- we try to do the exact same thing two different times. If you don't get the same output, then that's a real limit to what you can do, right? Because you've done everything you can. You said, here's the sequence of that DNA that has the instructions, this promoter sequence. It's exactly the same, yet you still get different outputs.

So, the question is, what's causing that? And also, how is it that life can actually function given this intrinsic noise that's in the cell? These are things that we're going

to look at over the course of the semester.

And these are actually some images that he took in this paper. And so, we can see that some of the cells are really rather red. Some are rather yellowish green. And so, this is telling us about the level of those proteins in individual cells.

So now, we have some notion. So, somehow, noise is important in these molecular scale gene expression patterns. Now, there is what I think is really quite a beautiful paper by Sunney Xie [INAUDIBLE] at Harvard in 2006, where he combined a single molecule fluorescence with live cell imaging in E. coli. And this allowed him to observe individual expression events within individual cells, where every time one of these proteins was expressed, he got a little yellow spot, corresponding to this equivalent of a yellow fluorescent protein.

And so, he was able to watch as real, live cells made individual proteins. And from that, he was able to say, I think, some very nice things about what it is that's causing noise, such as what we talked about in that repressilator, or in the other Elowitz paper. And a lot of it just has to with this idea that if you're talking about low number events or low numbers of molecules-- DNA typically present, only one or a few copies per cell-- then that means there's some inherent stochasticity. Because that piece of DNA-- it's either bound by one of these motors, this RNA polymerase that can make the RNA, or not. And that is intrinsically going to be a stochastic process. And that kind of dynamic can lead to substantial heterogeneity, or fluctuations, in expression of individual genes.

So, it's kind of at this stage of the course that we start to think maybe a little bit more about some of the global aspects of what it is that a cell is trying to do. And in particular, if a cell is trying to, for example, swim to get to higher concentrations of food, what are the fundamental limitations that cell faces? How does it know what is uphill, what's downhill? So, these are cases in which we have to really understand something about the role of diffusion in the ability of these small cells to move in their environment.

And for example, here is an illustration of the Reynolds number, which is telling you

something about the relative importance of viscous forces versus inertial forces on these different organisms. And some of the way-- for example, how an organism such as us can swim is just qualitatively different from how a microscopic organism, such as E. coli, can swim. So we'll try to understand how that plays out and, in particular, how it is that E. coli can move towards higher food sources. And there's a very clever way that bacteria have that allows them to have really robust functioning of this chemotaxis process within the cell. And I think this is a neat example of the gene networks coupling into a higher level behavior that allows cells to survive in really challenging environments.

Another manifestation, actually, of fluctuations is this idea of pattern formation. And this is actually experimental data of in vitro-- so, if you take proteins outside of the cell and you put them on a two dimensional membrane. Now, these are actually the proteins that are responsible for finding the center of the cell. So, I told that E. coli, for example-- it grows in length. And then, once it gets long enough, it wants to divide, so it separates in the middle.

And the question is, how does it know where the middle is? You know, if you can just stand outside a cell and look at it, then you say, I know where it is, and you just cut. But imagine you're a cell. How do you know where this-- once you start thinking about all these challenges that cells face, it's really remarkable that they can do anything.

And what it turns out, is that they implement-- they use what are called these Min proteins that display what seem to be the equivalent of what you might know of as Turing patterns in order to cause these oscillations in the cell that allows it to find where the center of the cell is. So, we'll talk about this and how these authors were able to visualize these beautiful traveling waves of proteins, where they successfully bind to the membrane, and then are ejected off of it. And this results in beautiful patterns that are traveling, as you saw.

So, this was-- I'd say that these topics are what you might call traditional systems biology, in the sense that these are all things that of physics branch of systems

biology were all thinking about for the first 10 years. And over the last five years, maybe, there has been a greater interest in trying to understand how these sorts of ideas and principles may be relevant for larger scale. And larger the sense of, instead of thinking maybe about genes as this fundamental unit-- they try to understand how genes interact to form this decision making process. Maybe instead, if you think about cells as somehow being that fundamental unit, how is it that cells come together to lead to interesting population level phenomena?

And so, we talk about both what you might call evolutionary systems biology. So, how is it that evolution within a population behaves? As well as ecological systems biology. What happens if you have more than one species, and how is that the kinds of ideas we talk about in the first half of the semester are relevant in these population level phenomena?

So, in the first example that we're going to give is actually another paper from Uri Alon's group, where he showed that there's a very fundamental sense in which cells, through the evolutionary process, are implementing a cost-benefit analysis. And the question that he asked here is that, if you take an E. coli cell and you put it in different concentrations of the sugar lactose. Now, the question is, how much of the enzyme responsible for digesting lactose-- how much of that enzyme should you make?

You might say, well, you should just make a lot of it. But he said, well, at some point, there's going to be a problem. Because if you make too much of it, then you're going to be spending all of your energies making this enzyme. Whereas, on the other hand, if you don't make enough, then you're not going to be able to get enough of this sugar. So, like always, there's this Goldilocks principle. You don't want too little; you don't want too much.

And what he showed is that if he evolved these E. coli populations over hundreds of generations in the laboratory, but at different concentrations of this sugar lactose, what he saw is that the concentration, or the level of expression of the enzymes required to make that, to break down that sugar-- it changes over time. So, if you

have a lot of the sugar, then you want to make a lot of the enzyme. If you have a small amount of that sugar, then you want to make less of the enzyme.

So, that all makes sense. But this is a case where he could really demonstrate it in the laboratory using these microbial populations. It's a very beautiful example of how simple ideas of cost-benefit really give you insight into the evolutionary process.

Now, I told you before that part of the reason that we have to consider the role of fluctuations, or noise, in, for example, cellular decision making, is because of the low numbers of molecules that are often involved. So, if you have a small number of proteins, or small numbers of DNA, then the process is intrinsically stochastic. Now, the question naturally arises, why is it that we might need to consider stochastic dynamics in the context of evolution?

Because if you think about, for example, an E. Coli population, even in a small test tube, you might have 1 billion cells there. So, 1 billion is a big number, right? Much larger than 1. So, it's tempting to conclude from that, that actually, all of this stochastic dynamics, fluctuations-- maybe it's just not relevant for evolution.

However, if you think about the evolutionary process, fundamentally, any time that you have a new mutant appear in the population that may be more fit, may be less fit-- but every new mutant in the population starts out as a single individual. It's a trivial statement, but it has deep implications, because it means that every evolutionary process goes through this regime where you have a small number of fluctuations. So, this has very clear locations in many different contexts, and we'll explore it over the course of the semester.

And despite the fact that that evolution is intrinsically, you might say, random, what's interesting are cases where that randomness somehow washes out. For example, we're going to talk about what I think is a beautiful paper by Roy Kishony's group, at Harvard Medical School, who showed that if you take a population and you put in some new environments, there are going to different mutations. Some of them are going to be really good; some of them are going to be not so good.

You can imagine that of all of these possible beneficial mutations, they describe some distribution. Whereas, if you asked-- this is the frequency, or the number, of mutations as a function of how good that mutation is, you might say, it should be some falling function. Because you just not are going to get as many mutations that are really just amazing as they are. They're kind of good. But it's not obvious whether the curve should be exponential, or maybe it looks like this. It could look like many different things.

What Roy's group showed here in this paper is that, actually, in some very reasonable situations, it doesn't actually matter what the underlying distribution might be. Because if you look at the distribution of mutations that actually fix or spread in the population, those actually all look kind of the same, in the sense that they're peaked around some value. So, there's some sense in which the random process of evolution leads to some patterns that are probably not so obvious. And on the flip side, what it means is that if you go and measure how good are the mutations that actually appear on the population, that actually tells you surprisingly little about what that underlying distribution is, in terms of the effects of the beneficial mutations. So, there's some way in which the details kind of wash out.

And I think this is fascinating because a major theme or major challenge in systems biology is we want understand how these underlying parts lead to some higher level function, but we don't always know which details of the interactions are important for leading to that higher level of organization. In some cases, they're very important, but in some cases, not. So, a challenge that we're going to face over and over again is trying to understand, what are the key features that are going to influence the dynamics of this higher level system? And this is, I think, an interesting example of how some features of that underlying distribution are important, and some are not.

So, another interesting analogy between evolution and some ideas from physics is this idea of a fitness landscape. So, just like an energy landscape tells you something about the dynamics of a system-- for example, you can say, a ball should roll down a hill. Similarly, if you think about evolution in the context of how fit an

organism is as a function of some different parameters, you can get what might be nontrivial structure.

So, this is some illustration of what's perhaps a nontrivial fitness landscape. Now, the height here is some notion of fitness. So, we could imagine this could be the ability of a bird to fly. In that case, maybe these two axes could be the length and the width of the wing. Now, the shape of this landscape tells you something about how evolution is constrained. Because if the landscape looks like this, then what that's saying is that you have to actually evolve, in this case, maybe a wider wing before you can evolve a longer wing. So, this is-- if there's structure to the fitness landscape like this, then it tells you something about the path of evolution.

Now, in this case, we're thinking about this in the context of phenotypes-- things that you can just look at the organism and measure. But instead of think about this in terms of phenotypes, we could instead think of it in terms of genotype. For example, there is a beautiful paper that we're going to read from Daniel Weinreich where he, in the context of a gene that encodes an enzyme that breaks down antibiotics such as penicillin, what he did is he made all possible combinations of five point mutations-- single limitations in the gene.

So, he made all 32 combinations of this gene and then measured the shape of the resulting fitness landscape from those 32 different versions of the gene. And from it, what he found is that there's a very interesting sense in which evolution, at the molecular scale, is somehow constrained. So, the idea there is that there somehow is a rugged fitness landscape that is constraining the path of evolution.

And so, in all these cases that we've been talking about, there's some notion that you can say, this organism has this fitness so long as it has a wing shape that looks like this. Now, that is perhaps find in many cases. But in some cases, there are what you might call game interactions between different organisms in a population. And what I mean by game interactions is that the fitness of a particular organism may depend upon what other organisms are out there. And in that case, you can't just say that one organism is fit or not, because it just depends on what everyone else in

the population is doing, or the genotype of the other individuals in the population.

So, in that case, you really have to apply some ideas from game theory to try to get insight into the evolutionary process. And we're going to talk about some really nice cases where researchers have constructed, for example, a rock, paper, scissors game using different E. coli strains. And if you stick out to that long, I'll tell you about a case in lizards where people have demonstrated a rock, paper, scissors interaction in the context of different mating strategies of the male lizards. So, if that's not an advertisement to stick around for a couple months, I don't know what is.

There are other cases where people have demonstrated that microbes interact via cooperative interactions, in which it's possible for cheater strategies to arise, spread throughout the population, and cause some harm to the population-- maybe even collapse of the population. So, this is a case where there's tension between what's good for the individual and what's good for the group.

Now, organisms are able to do a remarkable set of things. So, we saw cases where, for example, that neutrophil was able to chase the staph aureus, that bacterial cell. So, that's amazing. But that's responding to something that is an immediate part of the environment. So, it's chasing a bacterial cell.

But you might ask, is it possible for cells to learn? And, of course, then you have to define what you mean by learning. And there's been some really interesting demonstrations of how it's possible for organisms to learn not at the individual level, necessarily, but at the population level, via evolution.

And in particular, in this very well written paper, what they were able to demonstrate is that both yeast that have evolved in the context of wine fermentation and E. coli that have evolved traveling through, for example, our digestive tracts, there are characteristics, sequences of events, in which things happen. So, the idea is that if a bacterial cell is ingested by a mammal, they will typically see one carbon source, and then another one. So, they might see carbon source A, and then carbon source B, as they travel through the digestive tract.

But if that is typically what happens, then what it means is that an organism might have an advantage if, when it sees carbon source A, it starts preparing to digest carbon source B. So, we can actually learn something about typical environmental orderings. But it's not learning at the individual cell level. It's learning over the course of evolutionary time scales, the typical sequence of events. And in this paper, they show that this seems to be the case for both E. coli and for yeast in the context of fermentation. So, I think it's a really beautiful example of different notions of what you might mean by learning.

So, another classic debate within the field of evolutionary is this question of, why sex? And in particular, there's this classic paradox which is saying, sex is costly. In particular, if you take a bacterial cell, just one cell turns into two cells. And then two can turn into four, and you get very rapid exponential growth of the population.

Whereas, if you have both males and females, then there's what you might call this twofold cost of sex. Because males are, in some sense, not contributing to that exponential growth rate. If you start with a male, female, and they have two kids, and you have another male, female, and then you don't get exponential growth. And this is a factor of 2 in the rate of exponential growth. So this is what's in that exponent. So, this is a big, big effect. And so, I think it's a major, major challenge to ask, why is it that sex is so common among what you might call the higher organisms.

And there are many hypotheses. We'll talk about some of them. One of the leading ones is known as the Red Queen hypothesis from this Lewis Carroll story, where there's this line. The Red Queen has to run faster and faster in order to keep still where she is. That is exactly what you all are doing.

And the reason that it's called the Red Queen hypothesis is because it's arguing that perhaps the reason that we and other animals have obligate sexual reproduction is because of some arms race with parasites-- that the sexual reproduction allows us to evolve more rapidly against the always adapting parasite populations. And of course, we'll have to talk about exactly what this means. But

there have been some interesting experiments in worms, in which they had different reproductive, different sexual strategies, in the presence or absence of parasites. And this showed that there are some interesting cases where this may be at least part of what's going on.

So, at this stage, we've been talking first about decision making within cells, and then how evolution may allow cells to anticipate different environmental changes, may be able to play games against other strategies. But at the end, we're just going to talk some about inter-species interactions, and what these sorts of ideas may be able to say about that process. So, for example, a classic inter-species interaction are predator-prey interactions.

And this has been used to explain, for example, why it is that many natural populations oscillate over time. There are simple models of a predator and prey that lead to such oscillations. And just over the last 10 years, there have been some really fascinating experiments where, in the laboratory, they were able to take predator-prey, show that they oscillate.

But then, in this case, they saw some features that they weren't expecting. The oscillations were maybe much longer than they were anticipating. And instead of oscillating 90 degrees out of phase-- which is what you expect from standard predator-prey models-- instead they were oscillating 180 degrees out of phase. And I think that this is a good example of how quantitative experiments in the laboratory can actually say something about the classic models of predator-prey oscillations. They are over 100 years old.

But if you go, and you make quantitative measurements in the lab, you see, actually, in many cases, things are different. And then, you can ask, why? In this case, they went-- they did modeling, and they said, maybe it's because of evolution within the prey population. And once they were able to-- they had hypothesized that from modeling. And then they went, and they did experiments where they prevented that evolution, or they prevented-- they reduced the heterogeneity in the prey population. And then they were able to show that those two features disappeared.

So, I think it's really a beautiful example of the interplay that we always hope for, which is that you do theoretically motivated experiments, and experimentally motivated theory computation. Ideally, you go back and forth. And together, you can really learn more than you would ever be able to do just by doing one or the other.

We're also going to try to say something about the dynamics of populations in space. So, just like these spatial patterns that we talked about before, in the context of maybe gene networks, there are also dynamics of populations in space. For example, when populations expand into new territory, what does that mean about the evolutionary process? Once again, some very nice experiments have been done over the last decade to try to eliminate this process.

And in particular, one of the things that was found is that this process of genetic drift, the role of randomness in the evolutionary process, is somehow strongly enhanced in many of these experimenting populations. Because somehow the effective population size that quantifies maybe the strength of noise is somehow enhanced. Because it's not the entire population matters. It's just the population at the front of this expanding population that is somehow relevant. So, we'll explore how these ideas play out.

And so, towards the end of the class, we'll try to think about some real ecological phenomena. In particular, we're going to have one lecture where we talk about tipping points in populations. It's a theme that my group, for example, has been very excited about recently. So, this here is data from the Newfoundland cod fishery. And it's an example of how many natural populations can actually collapse suddenly and catastrophically in response to deteriorating environmental conditions.

Now, what's plotted here-- this is essentially the number of fish that are caught as a function of time. And you may not be able to read this, but over on the left, this is 1850, and this the modern day. So, this was a very productive fishery for hundreds of years and, actually, even for hundreds of years before this.

However, in the '60s and '70s, improved fishing technology led to a dramatic increase in the number of fish that were caught here. And that increase in fishing

led, in the early '90s, to a catastrophic collapse of that population. Similar things have occurred, for example, in the sardine fishery off the coast of Monterey, and many other populations. So, the question here is, how can we understand these tipping points in populations?

And this is a case where some of the ideas we studied early in the semester-- so, these cases of interactions-- can lead to sudden transitions. And for example, this early example of a toggle switch that we talked about at the beginning-- so, this case where, if you have interactions within the population, it can lead to alternative states. And it's the same basic dynamic here. If you have interactions within the population, then you can get alternative states-- maybe healthy and dead, or extinct, locally.

So, you can imagine if you have these alternative states due to interactions within the population, and if you start out in this healthy state, and you start pushing it, then it's going to-- the feedback loops there will maintain a state where it's alive, healthy. And then all of a sudden, when it's not able to counteract this deteriorating environment, all of a sudden it's going to switch and maybe collapse in this fishery. So, these sorts of ideas have been used to both try to understand why it is that populations might experience tipping points, but also to get some guidance about ways that we can anticipate that these tipping points are approaching?

For example, in my group, we've been excited about this idea that there are predictions that the fluctuations of the population should be different when a population is approaching one of these tipping points. And, at least in the laboratory, we were actually able to measure a change in the fluctuations before a collapse. And this is saying that, in principle, there are maybe universal signatures of populations and other complex systems before one of these tipping points.

Now, in the last lecture, we're going to go maybe to the largest scale to think about whole ecosystems. And this is, by its nature, I'd say, less experimental than the rest of the semester in that, in this case, we're trying to understand questions like, what is it that determines the abundance of different tree species on Barro Colorado

Island? So, it's an island in Panama where they go and they just count every tree within some region. They'd say, this is this country, this is that country. They count thousands and thousands of trees.

The question there is-- some species are more common than others. And we want to know why. It seems like a simple question. And the way that we normally think about this is if it's more common, then maybe it's because it's better adapted to that environment. And I think that, often, that's the right answer.

But there's been an interesting movement within ecology recently where it's been pointed out that many of the patterns that people have observed in terms of this relative species abundance-- how abundant some species are as compared to others-- that many of those patterns can actually be explained by a purely neutral model. I.e. This is a model in which you assume that all of the species are the same. So, this tree is just the same as that tree in terms of-- no tree is better than any other tree. But just because of the stochastic dynamics, random birth death processes, you can recover patterns that look an awful lot like the patterns are observed in nature.

So, you can interpret this is in multiple ways. But, of course, one way that we should always be thinking about these things is that if you observe-- we want to collect quantitative data. We should do that.

But there's always a temptation that if you collect quantitative data, and then you write down a model that is consistent with that data, we often take that as strong evidence that the assumptions of our model are correct. Even though we know that's not the way we're supposed to do science, somehow it's just really easy to fall into this trap. And I think that this particular example of this neutral theory of ecology is a very concrete example of how different models that make wildly different assumptions about the underlying dynamics-- they can all look the same once you look at a particular kind of pattern. And so, it's a nice cautionary tale saying, what is it that you can learn about the dynamics of a system or of a process based on a particular kind of data set.

And then, after this, we will just have that final exam. That's going to be that week of 15 to 19, I believe. So, once again, do not book your tickets before then.