Searching for simple models

Lecture Introduction

Trueswell
Hi, I’m John Trueswell, I’m the director of the Institute for Research in Cognitive Science. Welcome to the 9th annual Benjamin and Anne Pinkel Lecture. The Pinkel-endowed Lecture Series was established through a generous gift from Sheila Pinkel on behalf of the estate of her parents, Benjamin and Anne Pinkel. The series serves as a memorial tribute to their lives, to the lives of her parents. Benjamin Pinkel received a bachelor’s degree in electrical engineering from the University of Pennsylvania in 1930, and throughout his life he was actively interested in the philosophy of the mind, and published a monograph in 1992 on the subject, entitled “Consciousness, Matter, and Energy: the Emergence of Mind in Nature.” The objective of the book was, and I quote, “A reexamination of the mind-body problem in light of new scientific information.” The lecture series is intended to advance the discussion and rigorous study of the deep questions which engaged Dr. Pinkel’s investigations. Over the last nine years, the series has brought in some of the most interesting minds in the field of cognitive science as it pertains to thought, learning, and consciousness. These include Daniel Dennett, Liz Spelke, Martin Nowak, Stan Dehaene, Geoff Hinton, Ray Jackendoff, Colin Camerer, and Elissa Newport. So without any more description of this, I’m going to now hand the microphone over to Ben Backus, who will be introducing this year’s speaker, Professor William Bialek.
Thank you John. Have you all turned off your cell phones? OK. It is my very great pleasure to introduce to you this year’s Pinkel lecture, William Bialek. Professor Bialek is the perfect Pinkel, having been engaged with the rigorous study of deep question for many years now. He’s perhaps best known for his contributions to our understanding of coding and computation in the brain. He and his collaborators have shown that aspects of brain function can be described as essentially optimal strategies for adapting to the complex dynamics of the world, making the most of available signals in the face of the fundamental physical constraints and limitations that prevent that information from being measured and processed. One example that some of you will be familiar with is that 8 years ago, Professor Bialek coauthored a seminal book called “Spikes.” It is a classic in the field of biological information coding, and if you want a clear, rigorous, and intuitive introduction to statistical inference in the context of the nervous system, chapter two of “Spikes” is for you. Professor Bialek gave a talk here at Penn in the physics department in February of 2001. That talk was one of the clearest and most interesting talks any of us have ever heard. We can’t hold Professor Bialek to that standard at all times, but we certainly do look forward to what he’ll say today. He is the John Archibald Wheeler/Battelle Professor in Physics at Princeton University; he also has connections to the departments of molecular biology, applied and computational mathematics, biophysics, neuroscience, and quantitative and computational biology. He’s also a member of the multidisciplinary Lewis-Sigler Institute. He gets around. Professor Bialek went to public schools in San Francisco, then to the University of California at Berkeley, where he received an AB in biophysics in 1979, and a Ph.D. in biophysics in 1983. After post-doctoral work in the Netherlands and then doing theoretical physics at Santa Barbara, he joined the faculty at Berkeley. That was 1986. Then in 1990 he moved to the town of Princeton to join the newly formed NEC Research Institute. He joined the Princeton University faculty as a professor of physics in 2001. I’ll mention just two of Professor Bialek’s innovative teaching activities, because teaching and education have been so important to him. Professor Bialek was a co-director of a computational neuroscience course at the marine biological laboratories in Wood’s Hole, Massachusetts for a number of years, and more recently he’s experimented with education at Princeton to create a truly integrated and mathematically sophisticated introduction to the natural sciences for first-year college students. Now please come, if you can, to the panel discussion at the Institute for Research in Cognitive Science this afternoon from 3 to 4:30 PM, where you, Professor Bialek, and several of our distinguished Penn faculty from across the university will continue the discussion started by Dr. Pinkel now. I’m sorry, Dr. Bialek. Well, both, yes. That location, IRCS, is 3401 Walnut Street, A Wing, so you enter by the corner near Starbucks, and again that starts at 3 PM. Also there will be a reception outside of this room, following immediately after Professor Bialek’s talk. On behalf of Benjamin and Anne Pinkel, the Institute for Research in Cognitive Science, and the University of Pennsylvania, welcome.
Lecture: Searching for simple models

Bialek

So there’s a phenomenon with which I think many of you are familiar, which is the fear that you’ll be discovered as a fraud. This is apparently not uncommon among young academics. So I feel that a little bit giving a lecture in cognitive science. I’m a physicist by education and by culture, although my interests over the years have often focused on problems related to how the brain works and the phenomenology of the nervous system. I’ve had trouble in my own mind crossing this barrier between things that are a low-level function of the nervous system and things that we might dignify with the word “cognitive.” It seems there is a difference somewhere, although perhaps it’s not always so easy to articulate. So perhaps to confront those demons, I decided that I would do the best I could today, which is to, as I said in the abstract for those of you who saw it, talk about the search for simple models. So, in physics, we’re very familiar with the idea that you should certainly always try to do the simplest version of any problem that you’re interested in, and even if the simple version doesn’t capture everything, you might at least build your intuition in order to go on to more complex problems. Now I really don’t know whether that strategy will work in thinking about cognitive phenomena or not. It might be that there is, to borrow a phrase from the intelligent design folks, some irreducible complexity, sorry for the reference, that we just can’t get past, and so the notion that we’re going to build up our toolkit, as it were, our intellectual toolkit, by working on simpler problems, and then eventually we’ll get to things that we call cognitive, that might just be wrong. But I’m going to try it anyway, and I guess that the spirit of the lecture series is that it’s supposed to be exploratory and occasionally provocative, and in that sense I suppose I don’t need to apologize.

Anyhow, so we do a lot of things that we might put under the banner of cognitive phenomena, as you all know better than I do. Some of them are so obvious that you don’t even need to say what they are, (I don’t think, at one point when I was trying to make this slide, by the way, this isn’t, they’re not paying me, it just so happens that it’s a very convenient source. I presume that there’s a studio somewhere in Philadelphia where you can go visit and see some of these. Some of them are even comfortable.) I was making this slide, and I realized that I didn’t even need to write “chairs.” So obvious. So one thing we certainly do is to build categories that are useful to us and enormously flexible. The notion that we, without any difficulty, well, except perhaps for that one, or that one, call all of these things chairs, even if they’re designed by ((inaudible)), is quite remarkable. But of course we don’t just do that, we build big categories which are useful but then we also draw distinctions within those categories. Sometimes these are more subtle than others. If you look in the catalog, it’s actually claiming that that’s a dining room chair. We did the experiment at our house, it’s not a dining room chair. Anyhow, again, there’s this object. Look, I’m not telling you anything you don’t know. You understand the idea.

So the question is, can I take this problem of categorization, which is clearly central to how we think about the world, if we had to think about every view of every object separately we’d never get anywhere, can we find a simple example of this problem that
the nervous system has to solve? I’d like to suggest one, and that is to think about what happens when you see, not under the conditions you experience now, but under the much more limited conditions where it’s very dark outside. So that simplifies things, because of course then of course you only need to pay attention to your rods, and if you record the current flowing across the rod cell membrane, and you have a dim flash of light at time zero, you see a little pulse of current. You could say that it’s very dark outside and you were just trying to tell whether you were seeing something off in the distance. In a sense, this is the world you live in. These are the signals. They’re not these nice rich complicated things like the images of the chairs; they’re rather these little bits of raw data. Now, you might say, well, what’s the problem, this bump must be here because you saw something. The problem of course is that if you flash the light many times, you actually get all of these things. These are experiments from a former student of mine, Fred Rieke, who’s now at the University of Washington. In a sense, one way to think about the problem that you face in literally making sense out of this kind of data is that you would like to put these into categories. I mean, you have a sense that without too much discussion, well, I’m not sure what that is, but this looks like nothing happened, and that looks like nothing happened, and that looks like nothing happened. Some of these apparently were nothing. Then some of them look like something, and some of them look like bigger somethings, and maybe we can get somewhere here.

Let me, as an aside; one of the reasons why I think it’s useful to drop down to the very input stage of your sensory system, even if you think of yourself as a cognitive scientist, is to remember that these are the data that the nervous system actually has to work with. You don’t get to work with what the image actually looks like, because that’s constructed, and you don’t get to work with light intensity, because that’s the idealized physical stimulus. What you really get to work with is what the receptor cells do. And among other things, receptor cells are noisy, and apparently they’re variable even in response to repeated physical stimuli. So what’s going on here? Actually, this is not a mystery. I’ll remind you that in 1942, Hecht, Shlaer, and Pirenne taught us that when it’s very dark outside, the visual system is actually capable of counting essentially single photons. If I flash a dim light against a dark background, sometimes you see it and sometimes you don’t, and as a function of the light intensity, there’s some probability to see it. But it turns out that this probabilistic behavior in human subjects is to a large extent not the result of randomness in the brain, it’s actually randomness in the physics of light. That is to say, if you set the intensity on a light source, what you determine is not the amount of number of photons that arrive on your retina, but rather the mean number, and the actual number is poisson distributed. If you imagine the people that are willing to raise their hands and say “yes, I saw it,” when they count up to six photons, then you can predict what the distribution will actually look like, what the probability of seeing will look like as a function of intensity. This is a classic piece of work and it of course launched a large body of ideas: Why do you have to count up to six? Why is it that some people are willing to say yes when you they seem to count up to two? What does this have to do with the responses of the individual photoreceptors? And so on. It’s a remarkable piece of scientific history that I recommend to all of you.
So I know what these categories are supposed to be. These light flashes here are so dim that they’re delivering a small number of photons on average. And so the categories that the nervous system needs to build are zero, one, two photons. Very simple example of categorization. So you might say, well, it’s so simple that I know how to do it. Whenever a photon arrives I get a pulse. If two photons arrive I get a pulse that’s twice as big. So why don’t I look at the time when the pulse reaches its peak, and measure the amplitude, and figure out if it’s zero, one, or two? So you can try that. Look at the current that’s flowing across the rod cell membrane at the time of the peak response. This is the probability distribution. You can see that there’s a peak that corresponds to zero photons, a peak that corresponds to one photon, and a peak that corresponds to two photons. The only problem, of course, is that these peaks overlap tremendously. In here, or in here, you’d be quite confused about what you were seeing, how many photons you were seeing.

Actually, you know, it’s difficult. I mean, these are recordings from a salamander rod, you can do the same thing in monkeys, but if you allow me a little bit of license here, I think we can claim that the Hecht, Shlaer, Pirenne experiments and their generalizations, in which you actually ask people how did you think bright the light was and so on, essentially would be inconsistent with this level of ambiguity. This is telling you, basically, count them all, and this is saying you really can’t tell if you have zero or one half the time. So what’s going on here?

Well, as I said, this is a problem of categorization. The data that you’re using to categorize, unlike the images of the chairs, if you will; what are the images of the chairs? They’re little things that are, what, ten thousand, twenty thousand pixels, so there’s twenty thousand numbers, times three for color, that go into your categorization of those images. So you might say, this is simple, right, it’s just one photoreceptor, so there’s only one number. But that’s not true. Because of course when one photon arrives it produces a current pulse in time. So actually there are many dimensions to this space of data that you’re taking in. There’s the whole sequence of current that you observe across time. This simple categorization that we tried just says, well I’m going to look at one moment in time, so I’m going to take this nice long list of data, and I’m in this multidimensional space, I’m just going to look along one axis. It’d be like looking at once pixel of the image of the chairs, and you know that wouldn’t work, in fact nobody would even suggest that. But in this context, it seemed like a good suggestion, and it wasn’t bad. But of course you can do better. You can ask, how should I really divide up this multidimensional space into categories, is there a better way of doing it? The answer is that yes, there is a better way of doing it, and the way of doing it is sort of buried in the data itself. That is to say, some of the data consists of signal, if you look at the average response, you can see that the shape of the waveform in response to a single photon is pretty stereotyped, and then there’s some noise rumbling in the background. So really what you want to do is to find a way of picking out the coordinates in this space where you get as much of the signal as you can and as little of the noise. If, in order to do that, all you need to do is to look in the right direction in space (it might be more complicated than that), so instead of looking along one axis which corresponds to a single point in time, you look at some linear combination of different points in time, which of course is
some other direction in space, then what you’re really saying is that I should take this current as a function of time, and filter it. If I build the right linear filter, I can in fact remove almost all the ambiguity. So the dashed line is what we had before, and the solid line is the output of the best filter I can build. This particular data set isn’t so big, so there are still a few wiggles, but you can see that the troughs in between zero and one and one and two now almost reach zero, which basically means that you can draw a line here and the little bits of ambiguity have now shrunk to be of order a few percent.

If that’s really the answer, then maybe that’s actually what the visual system does. If we take seriously this notion that you should build the filter that gives you the best separation of signal and noise, then a very important fact about that is that that filter is determined by the signal and noise properties of the photoreceptors themselves, not by anything else. In a sense what you’re saying is, I think that one thing that the visual system needs to do, is to build a filter that’s matched to the characteristics of its own receptor cells. So if I tell you all I can about the receptor cells, I should be able to build that filter for you, and if you think about it a little harder, you realize that in order to make this work, you should put that filter essentially right behind the photoreceptor, or right behind the photoreceptors of the bipolar cells. If you predict what that filter should be, what the output of that filter should look like, in fact, that is the output of the bipolar cells. No preparameters. I’ve shown here the voltage response of the rods. As many of you know, the dynamics of the cell membrane impedance mean that there’s a nontrivial transformation between current and voltage, and you might have worried that all those filters are actually implemented in that transformation, and it’s not. You can see the difference, and you can see the quality of the agreement, and so on. This is an example of a very simple version of a problem that the nervous system has to solve more generally. What’s interesting about it, I think, is that because it’s so simple, we could push to the point of saying, well, how well does the system really solve it? And if it solves it well enough that it becomes interesting to ask, how would you solve it optimally, then the question of how you would solve it optimally becomes very well posed mathematically. Now, the question of how the retina actually implements this filter in molecular mechanisms, that of course I can’t tell you anything about by this argument. But I can tell you what it should compute.

So that’s an idea. Now, will that generalize and so on? Let’s keep going. The example I just gave you, I motivated it with the chairs, and let’s be honest, that is at best an analogy. Right? Yes, in some way you’re categorizing these images, and yes, you’re categorizing the photoreceptor currents, but it might be that those problems are so different from each other in practice that saying that they’re similar conceptually doesn’t help. I don’t know. But sometimes it’s not just an analogy. Sometimes, very “simple” creatures and very small parts of the nervous system solve problems that we solve. And it’s not just that the problem is kind of similar, it’s the same problem. So if you watch my hand move, you follow it with your eyes, and in order to do that you have to have an estimate of its trajectory, and you have to know how fast it’s moving. And so among other things, we use our visual system to estimate motion. In fact, we use it not just to follow objects with our eyes, but to know where to reach for them, we use it to know that when we’re walking, the pattern of visual motion tells us that we’re walking straight, as
opposed to wobbling around. Our friend here in fact does the same thing. She takes the signals from her visual system and estimates motion and, among other things, uses that signal to stabilize her flight, which is a nontrivial problem. I’m not going to go through the evidence that that’s true. Let me, however, make the observation that the transformation, the fact that there are in the fly’s visual system neurons that are sensitive to motion is true, and they’re conveniently located near the back of the head. Doesn’t make the problem any easier. We have neurons in our visual system that are sensitive to motion, in fact we’ve got a whole neighborhood in the visual cortex called MT, or V5 depending on where you live, that does this, and of course it’s a computation. The data that come in through your retina don’t know about motion, the photoreceptors don’t know about motion, what they track is light intensity as a function of space and time, at best.

Before I go too far in talking about flies, let me dispel an apparently commonly-held myth, that the fly’s visual system is very different from ours because it’s a compound eye and has lots of lenses. In our eye, as you all know, in the back of the eye you have the photoreceptors sitting in the retina, and they’re looking out this way, and there’s one big lens on this side. So the photoreceptor that’s sitting up here looks out through the lens like this, and the photoreceptor down here looks out through the lens like that. Suppose that instead of having the photoreceptors like this with a lens over here, you put a lens in front of every photoreceptor on this side. In that case, this photoreceptor would look out through its own private little lens in that direction and this photoreceptor would look out through its own private little lens in that direction. That’s the way the fly’s eye is built. That’s the difference between a compound eye and a simple eye. The big difference is spatial resolution because diffraction depends on the size of the aperture that you look through. The spatial resolution in looking through a big lens is much better than the spatial resolution in looking through little lenses. So if you’re an insect, and you have a compound eye, your spatial resolution is about a degree, which is the width of your thumb at arm’s length, which is the width of a face seen across the room, so it really is a very blurry view of the world.

The good news, of course, is that you don’t have all that wasted space between the photoreceptors and the lens. That is, after all, a space that is large enough, if you’ll forgive the image, to fit the entire fly. For a big animal like us, this is a perfect sensible strategy, for him it would be very bad. Correspondingly, if we tried to achieve the same spatial resolution with a compound eye, we’d need a head that was meters across, which even academics don’t achieve with age. This is actually kind of a historic document, it’s a case in which Mr. Larson got the biology wrong. He doesn’t usually. Each photoreceptor is looking out through its own individual lens (that’s not exactly true, there are actually 8 photoreceptors looking out through one lens), but the point is each lens is responsible for seeing a small fraction of the world. So therefore the lenses here are much more like the pixels of a camera than what would be indicated here. Although it is kind of charming.

The fly uses its visual system to compute motion, and it uses that to guide its flight. As an aside, I would caution you against the popular belief that this is a hard-wired reflex,
since after all it’s an insect and it doesn’t learn; all of those statements are false. The reason that fly’s transformation from visual input to motor output is not a hard-wired reflex is not biology, it’s physics. You see, if you suppose that your visual system told you that you’re rotating 20 degrees per second to the left, what do you do? You have to beat your wings differently. Well, how much should you beat your wings differently? You have to produce a torque that pulls you back. Actually, how much torque should you produce? Well, it depends on your mechanics, which by the way depend on what you just ate, if you’re a fly, because they can change their body weight noticeably by eating. Furthermore, the mechanics of their wings are actually quite complicated. They fly in a regime of intermediate Reynolds number, where things are very unstable. And actually the geometry of their wings is not encoded in their genome, because of course they undergo metamorphosis, and the wings harden when they emerge from the pupa. So the transformation from visual input to motor output must be learned, because the physical constraints can’t be encoded into the genome. In fact, the ability of flies to use visual input to guide their motor output as they fly around, is at least to some extent a learned ability, which surprises some people.

There’s a beautiful experiment by Martin Heisenberg. There’s a history to this. It’s really off the point but it’s such a good story. One of the classic experiments is to take a fly and hang it from a torsion balance. So it can’t actually rotate, but you can measure the torque that it produces. Then you blow on it a little bit so it starts flying. They won’t all do that, but many of them will; flying, that is to say, beating its wings, it’s not going anywhere. Then you give it different visual stimuli, and you show that, for example if the world is rotating, then it tries to produce a torque to counter-rotate to give itself the impression that it’s flying straight. So then, of course, what you can do is to take the torque that it produces and feed it back to the visual stimulus, and now you’ve built a flight simulator for a fly. Of course, there’s a parameter here, which is how much torque produces how much rotation, and you can even think about something more complicated in which the torque moves an object that’s in the front of his field of view while the thing in the back stays fixed, or something like that. Actually you can do yet more sophisticated versions, in which the fly is standing on something and when he pushes with his right foot, it rotates one way, and when he pushes with his left it rotates the other way. The paper which describes all this, of course, is called “Can a Fly Ride a Bicycle?” for obvious reasons. This is a sensory motor association that actually doesn’t occur in the natural history of flies. It’s not true that if they apply more force to the right foot the world rotates, that’s not what’s going on. Nonetheless they will learn to stabilize an object in front of them. The really impressive thing is that you can switch the sign and then they learn that too. More than meets the eye, as they say.

Let’s focus on the part about motion estimation. As I said, in order to fly, they use visual signals, they extract information about motion, and then that appears in various neurons in the back of their head. I spent a long time collaborating with a good friend, Rob (de Ruyter van Steveninich), who’s now at Indiana University, the physics department there, and over the years we tried to poke around in this little corner of the fly’s visual brain to see if we could understand something. The latest incarnations of those experiments, actually, are quite beautiful. One of the threads of those experiments is trying to explore
how the nervous system deals with ever more complex and naturalistic stimuli. At some point Rob got tired of trying to generate those stimuli in the laboratory, so what he did was he miniaturized the electronics for recording from the fly. The fly’s actually here. In the years that we spent together at NEC, it was nice because we could do things related to how the brain worked, but it also looks like a physics lab, with racks of electronics and so on, and every once in a while if you looked closely you could see that there was an animal. Anyhow, here, what they’ve done is to miniaturize the electronics so that this whole thing can actually be mounted on a motor drive, and you can take the whole thing outside and literally fly the fly around and record its motion-sensitive visual neurons while you’re doing this, so you get some understanding of what these neurons do with the real-world signals that you actually encounter. As a cautionary tale, for many things in neuroscience, they do things that you definitely wouldn’t have expected from experiments in the laboratory. In particular, it was widely believed that these neurons were incapable, they have tuning curves for velocity, which is interesting because there aren’t cells tuned to different velocities, which makes me wonder whether your interpretation of the tuning curves is correct, and if you really thought about it, it might make you wonder whether your interpretation of the turning curves in brains like ours where there are neurons with multiple tuned velocities are correct. The tuning curves don’t actually look all that different.

Anyhow, the tuning curves that you measure in the laboratory, you sit them down in front of a television screen and you move a sine wave back and forth, they cut out at a hundred degrees per second. When flies are flying, they can do acrobatics at 2,000, 3,000 degrees per second, and so the conclusion was that these neurons play a role in guiding those motions. Well, take the neurons outside, where the lights are brighter, the contrast is higher, and you’ll notice that the fly’s eyes go all the way back around his head, and so the area of the retina that is being stimulated is an order of magnitude larger when you’re moving around in the world than if you’re looking at a little TV screen. All of these factors together result in these neurons being able to follow velocities of 3,000 degrees per second with no problem. There are things to be learned even in very simple terms by trying to take the nervous system into a regime more like the one it encounters in nature, and that was the point of this setup.

In order to do all this, the fly has to do at least two things. One is that it has to take the data that it collects in the retina and turn it into an estimate of motion. The other thing it of course has to do is to write down the answer, as it were, everywhere in the nervous system in a sequence of action potential. You could also say it has to divide up the signal among many neurons, that population aspect of this system is relatively simple and if you think of one neuron at a time you’re probably OK. I’m actually going to focus on this problem of motion estimation, which means unfortunately that what I’m talking about is how you extract the feature that you’re interested in, not how you build a representation of it once you’ve done it. That, in retrospect, may be a strategic error for this audience, but having made it, I’ll keep going. If somebody’s curious about the representational problem we can talk about it later, and we’ll touch something that has more of that flavor later.
Of course, in the spirit of the idea of optimal filtering that we talked about before, you might wonder how accurate these estimates really are, and whether they’re as accurate as they can be. In order to address this question, you could think of a variety of ways of doing it, what we tried to do some years ago was to just ask, if we look at the action potentials coming out of these neurons, and you should imagine here that there are two neurons, there’s one on the right-hand side and one on the left-hand side, and they’re arranged with opposite direction selectivity, so one’s being excited and one’s being inhibited, could we look at this sequence of action potentials, and reconstruct what the trajectory was that the fly actually experienced? For a variety of theoretical reasons, we thought to try something very simple, which is essentially linear filtering, in which every time you see a spike it means something and you just add up the results. Somewhat surprisingly, that actually works. As I said, we had theoretical reasons for believing it ought to work, but there’s theory and there’s experiment. Check.

So having done that, basically you do the best you can to decode the signals, and of course it could be that there’s something hiding in there that we don’t understand, but this is the best we can do at the moment, so let’s quantify how well we did. The way to do that is to measure your errors and compute the power spectrum of your errors. The power spectrum of the signal, here expressed not as a velocity but as a displacement, looks like this. The power spectrum of your errors looks like this. At intermediate frequencies, you have a pretty healthy signal-to-noise ratio, factor of ten, twenty, thirty, something like that, and you hold that up at high frequencies, you’re holding the noise level somewhere around here, ten to the minus four degrees squared per hertz. By the way, if you translate this into how big a displacement can you detect in, let’s say, 30 milliseconds, which is how fast the fly can actually do these things, the answer is one-tenth of the size of the spacing between the photoreceptors, which means that these neurons are capable of delivering signals which are in the hyper acuity range. Actually, historically, some of the recordings from these cells were among the first examples of seeing single neurons performing in the hyper acuity range. So I’ll remind you, for those of you who don’t know, that we are also capable of detecting displacements which are much smaller than the space between photoreceptors on our retina, or, alternatively, smaller than the width of the diffraction spot as seen through the lens, and that’s a classic phenomena called hyper acuity, and flies can do it too.

What’s interesting is that in flies, we actually know enough about the physical limits under which the system operates that we can calculate how well you could do if you made use of all of the available signal-to-noise ratio. It’s hard to see motion when images are blurry, because it’s easy to see a sharp edge move, it’s hard to see something shallow move. The images are very blurry because of diffraction through these little lenses. In addition, the photoreceptors are noisy. Much of that noise is not the photoreceptor’s fault, it’s that it’s only counting photons at some rate, but you can measure the photoreceptor noise under the same conditions as these experiments. Actually, one of the reasons why we thought the fly was such an attractive system is that in this case the visual motion computation takes as input the data from the photoreceptors. You can basically spend all afternoon with a photoreceptor if you want to, and characterize its signal-noise properties as completely as you can for any cell. The
output of these computations, after four or five layers of processing, depending on how you count, is encoded in a small number of very large neurons which are conveniently located in the back of the fly’s head, where you can make a very small hole, which, insect exoskeletons being what they are, actually closes up and holds your electrode in place. You can record from those neurons for 5 days, which is a noticeable fraction of the life of a fly.

So you have this fantastic opportunity to study what is a simple visual computation, but also that we do, in a setting where you have complete ability to characterize the signals that you start with at the input of the computation, and you have very stable access to the neurons that, as it were, write down the answer to this computation. Among other things, you see that at high frequencies, which is actually what matters for the fly, because of course he’s doing all this while he’s turning pretty fast, the noise level in our reconstruction of the stimulus and the noise level that is the minimum determined by diffraction blur and photoreceptor noise, are very close to each other. If we wanted to work really hard, we could start discussing whether the missing factor of two was real or not. One of the reasons I’m not too worried about the missing factor of two is that, I’ll remind you, under these conditions, there are about three thousand photoreceptors being stimulated across the whole retina, which means that this limiting noise level has in it a factor of one over three thousand as you average over all the data. A thousand is from there to there. You sometimes have the idea that the solution to the problem of noise in the nervous system is, don’t worry, there are lots of cells, average. That’s true. It helps. But there’s still a limit. And as I said, this corresponds to a tenth of a degree on a behavioral playing scale, just to give you a flavor. So it looks like the fly really does get close to doing optimal estimation. So let’s say we take this seriously. You’ll understand what you’re looking at in just a moment.

There’s a mathematical problem here. I take the photoreceptor signals as inputs, and when I’m done, I’m supposed to get an estimate of the velocity. Now, if you’re willing to grant me a few hypotheses about the statistical structure of the visual world, which by the way we can now do better on, because Rob has designed an experiment that actually goes out and measure these things, it’s a really interesting piece of work, not yet done, but if you grant me those then we can set up this mathematical problem and in certain limiting cases we can actually solve it. I mean it’s a, I mean, it was a decent PhD in theoretical physics, so I’m not going to take you through all of the calculations, but let me emphasize that there are a couple of extremely limiting cases which are intuitively understandable. So for one, suppose the synchronizations are really high; lights are bright, contrast is high, and you’re not moving too fast. Well that in that case, you all know how to estimate velocity—go to a point in the image, take the derivative in time, take the derivative in space and take the ratio, and that will show you how fast it’s moving. And that’s right. That really is any reasonable theory of optimal estimation has to recover that as the high single noise-ratio limit. The question is, how do you approach it? And actually, we can tell you something about how you approach it. What happens at low single-noise ratios? The lights aren’t so bright, contrast is dim, velocities are either so fast that they add extra blur or so low that they’re hard to measure. In that case, it turns out that you really don’t want to take ratio’s of derivatives; all the experimentalists in the
audience know that when you take data and you take the derivative, you get more noise, so you should never differentiate your data if it’s noisy, and if, per chance, you absolutely had to differentiate your data, you shouldn’t divide by it, because you might divide by zero by mistake. So that can’t be the right answer at low single-to-noise ratio, and low single-noise ratio, as it turns out, the answer is that, essentially, you know something’s moving, because what you see at one point in space and time is correlated with what you see at a corresponding point in space and time; it’s linked along the x=bt line. So what you look for is just those correlations; you don’t actually try to follow an object. Let me emphasize that, whenever you set up a theory for doing optimal estimation, you’re always making a trade-off, and the trade-off is between insulating yourself against noise, that is, the random errors you may make, and doing things that are veridical. So, this, if there’s no noise, gives you exactly the right answer. The only problem is that it’s infinitely sensitive to the noise. This is very robust against the noise, but, of course, even on average doesn’t give you the right answer, because if I simply make the image twice as bright, the voltages of all the photo receptors go up by a factor of two and my estimate changes by a factor of four, and so I confound velocity with contrast, which, by the way, you do at very low contrast, and so does the fly. So this is all very interesting, but how do you test this? That’s a long story; it took us ten years to convince ourselves that we’ve seen all of the signatures predicted by the theory. I don’t know whether we convinced anybody else, but it took ten years to convince ourselves. But what I wanted to do to sort of tie-back the idea, that what we’re talking about is a computation that really is the same computational problem for us and the fly, is to show you a stimulus that was designed by Rob to probe aspects of these predicted mechanisms, and has the property there’s nothing actually moving, but if you compute motion in this way, you will get a non-zero answer for the velocity. So what you’re seeing is hopefully something that looks like it’s moving, to the right, and then it slows down, and then there’s nothing moving, and then it moves to the left. Let me assure you that there are no moving objects here; what you’re actually seeing is two screens, each of which is undergoing completely random flicker. Every square on the checkerboard is updated on every frame with a new Galician-random number for the light intensity. Then what we do, is, we look at them for a little while like this and then put them on top of each other. Now it turns out that, if you just did that, nothing would happen. But what’s special, right, you just have two random things that are on top of each other, you add them up, and you get random thing, right? But what’s special here, is that the two random movies are actually the same, except one of them is one frame ahead of the other. So you have two signals which are each completely random light noise, but one is a delayed version of the other, so it’s shifted in time, and then we put them on top of each other with some shift in space. And when you add them, that produces precisely the kind of spatial-temporal correlation that you should interpret as motion is you’re doing this computation, or even if you’re doing this one actually. Technicality. And you do. Unambiguously. And what’s interesting, of course, is so does the fly. In fact, it’s generalizations of this kind of stimulus that have led us to think that we can actually dissect the form of the computation. So the message here, is, the fly is doing a problem that seems, on the surface, to be very much like a problem that we have to solve; it solves it very, very well - close enough to be an optimal that it makes you think you should explore the theory of optimal estimation as a theory for what the fly actually computes. If it does that computation, then it should experience certain illusions,
and not only does the fly experience those illusions in the sense that the motion-sensitive
neurons respond, but we experience them too, which suggests, perhaps, that we’re doing
computation that’s not so different.

I’ve talked to you so far about things in which there’s one photo-receptor, or, well, many
photo receptors, but it funnels down to one neuron. But of course most of the interesting
things that the nervous system does involve lots and lots of neurons. And as many of you
know, technology has improved to the point that we can record from lots and lots of
neurons. This particular image is from my colleague, Michael Barry, and is a slice of a
salamander retina. This is the same salamander that the rod cells came from before; it
doesn’t really matter, but he’s kind of cute. What you see here are, in the green blobs, are
retinal-ganglion cell bodies. The streaks are where the axons have bundled together and
will eventually form the optic nerve, and the black dots are electrodes on a glass slide that
you can use to record from these cells. And what’s important to take away from this
picture, and would be even more true of even more modern incarnations of this, this is
already one point five to two generations old, is that the number of black dots and the
number of green dots are approximately the same, which means that, with modern
fabrication technologies, you can pack enough electrodes in that you can basically get
one electrode per a cell, and, unless there’s a really terrible conspiracy, you should be
able to sort them all out and record, essentially, from all over the cells from this
neighborhood. So let me get you thinking about this, and I want you to think, a lot of
questions people ask about populations of neurons are in some sense the natural
generalizations to the population of a question you could have asked about a single
neuron. What I’d like you to think about are questions that are really only properties of
the population. So for example, let me slice time up into little windows, like in the frames
of a movie, ten or twenty milliseconds in this case, and if that ten- or twenty-
milliseconds is short enough, than in each window, either a cell spikes or it doesn’t, so if
I record form ten neurons, what I see, in a sense, in that little frames of the movie that the
retina is translating to the brain, is a binary word. And, of course, if there are ten cells,
there are a thousand and twenty-four possible words. But even in the salamander, the
number of neurons that are responsible for dealing with a certain small patch of world is
not ten; it’s more like a hundred. I don’t know why I did this, but I actually worked out
what two to the one hundred is, but you get the point, right? There is this combinatorial
explosion of the number of patterns that the retina can generate. I guess if it did all of
these things, the brain would never make any sense out of it, and one presumes that
there’s something orderly about the way in which it uses all these combinations. So let’s
try and explore that, and I think I don’t need to emphasize that although we’re going to
talk about the retina here, this is obviously a problem that you can ask about any part of
the brain, and it gets harder as you start thinking about the cortex, where if you just think
about one cortical column; it already has tens of thousands of neurons in it, so how do
you think about the states of that network? Theorists, we don’t contribute very much, but
we do make a few important observations, one is that, in some sense, if everything is as
complicated as it could be, you’re screwed, right? I mean, if you are looking at a system
that is built by the parts, and the parts can be put together in all possible ways, with no
constraints, right, then there’s nothing—what are you going to do—there’s no
understanding you an generate that goes beyond the observation that it’s got lots of parts
that can be put together in all possible ways, right, there’s nothing more to say. So the only reason you can say anything about a complicated system is because it’s not actually as complicated as it could be, or could have been. Now this is actually a general remark about the progress of science or not, I’ll leave that one for the philosophers, but in this case, it’s quite practical, because if you have n-neurons, which have two to the nth states, you are not going to do an experiment who’s duration is ordered two to the nth. So you’re going to have to try to understand what this network of neurons is doing, even though you record from it for a time, which is nowhere near enough for it to explore its full-face space. So what do you try? Well, suppose you try the idea that basically every neuron is doing its own thing. And actually, although that sounds silly, it’s really seen as a good approximation, because if you record for two neurons and you ask how correlated they are, so, you know, veritably the spiked didn’t spike and you just measure the correlation coefficient between two neurons, if you see a ten percent correlation, that’s big. That’s the edge of the distribution. So you might say, well, alright, maybe they’re….. And by the way, this is a quite common observation; it’s not just from the retina, right? Stick two electrodes in the brain, two cells near each other a plausibly involved in the same process, actually measuring correlations between their spiking isn’t so easy. They tend to be significant, but weak. Unfortunately, if you try the independent approximation not on two cells, but at forty cells, and you just ask “How many of those forty cells spiked a the same time?” , then, it turns out, if they were independent, you would predict that that probability would fall off very rapidly, but in fact, it falls off very slowly, roughly exponentially, as opposed to this kind of binomial-regiousian thing, so that there’s actually significant probability for ten of the forty cells that fire together, they do it one in a thousand times, but remember, we’re talking about ten milliseconds, so that means every ten seconds that happens, and you would predict, basically, that that would never happen. Ok, so this hypothesis, while it looked good, was actually wrong about the whole population. So let’s try another simplifying hypothesis. The reason ten cells fired together is because there’s something special about those ten cells and they’re connected together, or they share common input from amicrine cells, whatever. Unfortunately, this isn’t actually a simplification, because there’s lots of groups of ten cells, in fact, there are forty-two ten of them, so if you go around saying that every time you seen ten things do something together, that must be special, you’re not very much better than the guy that was trying to do the experiment that was two to the n wrong.

So, here’s another idea, that actually was captured nicely by our friends at our local rag, I presume you guys have one too, and that is that the neurons interact with each other, they’re not completely independent, but instead of acting in arbitrarily complicated ways, they only interact in pairs, but remember, there are lots of pairs, right, there are forty squares of them in a group of forty, half of that, but never mind. And the idea would be that the whole system can do something collective, for example, ten cells can decide to fire together, not because there was something special about those ten cells, but that was the outcome, if you will, of all these paralyzed discussion. Our friends sort of noted, that, you could sort of think about social analogues with this, where all of your opinions are shaped by conversation with one other individual, all the paralyzed conversations happen, and then you go out and do something. And you might, none the less, all decide to go do the same thing. Not because anyone gathered you in a room and said “All forty of you
should do the same thing”, but rather because you agreed upon it by sort of iteratively having all the paralyzed conversations. So we all know that collective action can emerge from paralyzed interactions. It also turns out that this is a very well known set of problems in statistical mechanics, as we’ll see in just a moment. So essentially, what this says, is, outside the illustration of course, is that the paper had four authors, and you’ll notice that there are only paralyzed discussions going on. Of course, the authors have to appear many times in order to have all the paralyzed concessions. Some of the paralyzed discussions presumably correspond to positive interactions, some with negative interactions, and so on. So what we’d like to do, then, is to take seriously this idea that things happen in pairs. Let’s take these very weak correlations between the pairs, and remember that since there are many of them, it’s possible for them to add up to something significant. What I’d like to do is build a model that captures these paralyzed correlations, but is otherwise as little structured as possible; I don’t want to assume anything else. So formally, what this means it that I’m going to write down a model which is the probability that this network finds itself in some particular state; that probability distribution has to exactly match these correlations, but otherwise should have as little structure a possible, and technically, “as little structure as possible” means “as random as possible”, so as much entropy to the probability distribution as I can possibly have.

So it turns out, you know, this was not lost on us when we started thinking about this, that the problem that I just posed, which is to build a model that is as random as possible but captures the paralyzed correlations between neurons that either spike or don’t, is exactly the Eising model, which, you know, warms the hearts of many physicists, maybe doesn’t do much for the rest of you. But what’s important about it is that we know a lot about this class of models. So before I tell you anything about what we know from the class of models, is this actually correct? That is to say, is this a good model of this network or not? So how do we test that? What we can do is restrict our attention to ten neurons, and that’s good because ten neurons is big enough that there’s lots of hairs, but two to the tenth isn’t that big, so that you can actually do an experiment that’s long enough to sample the whole probability distribution. And furthermore, you can do it many times, because out of forty neurons, I can choose ten in many different ways. So here’s typical results; let’s actually ask, at what rate do we see any particular binary pattern, and the rates of course range from very often, you know, a couple times a second, down to once per a hour, and the whole experiments only one hour long, so it’s really not very well measured. If you pretended that all the neurons were independent of each other, you get these blue dots, which are interesting; they’re actually anti-correlated with the data, and if you build this model that keeps track of the pairs, then you find that everything works, except down here where we’re not measuring very accurately. So what we would claim is that, in fact, by taking seriously the paralyzed correlation, even though they’re very small, if you add them all up correctly and build the simplest model you can that’s consistent with them, you none the less correctly predict the distribution of states that a network of ten neurons takes on, where as if you’d ignored the paralyzed correlation, you can make errors which are many orders of magnitude in probability. So in this sense, the network of ten neurons is very far from being independent, right, things happen at rates which are five, six magnitudes away from what you would expect, and we actually get
those right, just by keeping track of the paralyzed interactions. But this model has many properties. If you think of it as being a physical system, then this is the fulsome of distribution from statistical mechanics and thermal dynamics, and what’s up in the exponential here is the energy of this analogous physical system, except in this case it’s not an analogy, its-, this is the probability distribution. This energy can have many local minima, which correspond, if you will, to kind of stored patterns in the network.

[lecture tape ends here and does not currently pick up again till the last paragraph]

Networks somehow have stored patterns in it. Well, it turns out that if you work out what happens with all forty neurons, it really does. And furthermore, you can ask, you know, in the Hopfield picture of memory, right, you have this network and it’s got these states that are attractors, right, ground states of the model. So if you imagine taking an initial state ad rolling the ball downhill and energy, it will drop into one of the attractors, and those are our ground states or stored patterns. So what we can do since we now know the energy function, because we’ve built this model that describes the divide, we can ask, suppose we do that memory computation, we do the memory recall of the Hopfield model, and we take every pattern of spiking and silence that we see, and we figure out which attractor goes with it. And then we show the movie again. And again, and again. And an amazing thing happens, which is that you visit the same attractor over and over and over again, even though the microscopic states, that is to say, exactly which neurons are firing, is different every time. So these ground states that are sort of imbedded in the model, which came out of just analyzing the statistics, turn out to be things that the retina really uses in responding to movies, because, when you play the same movie over again, it spits out the same attractor, but not the same microscopic state. Which, by the way, is interesting right because, as an example, if you’re willing to let me stretch a little bit, it’s a little bit of sort of syntax versus semantics, right? When I just look at the states of the system, I’m doing syntax right, I’m just asking about the probabilities of finding things. When I play the same movie over again, I’m asking about semantics, does this pattern mean something in relation to what you’re actually looking at, and in this case, I can actually go from the syntax to the semantics, which, I understand, is something people worry about. Which might mean it’s a bad analogy. There’s more to say here; let me say that you can now ask is whether these models- we find these models and they do all sorts of interesting things, but the real power is in your ability to talk about larger and larger groups of neurons. And so we’ve been able to do that, even to go beyond what we’ve been able to measure experimentally in Michael’s group, and what we find, quite startlingly, is that, when we start to look at much larger networks, it looks as if the correlations in the system are poised at this very strange place, where if they were any stronger, than in some sense- I’m sort of translating loosely from what can’t actually be made mathematically precise so, I hope I do this right- so, if you made the correlations any stronger, the system would sort of just start to lock up. Basically, you have so many pairs that are correlated with each other, and the correlations get so strong, that you start to dramatically restrict the number of states that are accessible to the system, and it would be like a thermodynamic transmission freezing. On the other hand, if the correlations were weaker, you’d start to go towards this regime where every neuron is doing its own thing, and there’s a well-defined boundary, like face-transition single-mechanic, and the
system seems to be poised just about there, and if this peak were right there, then you’d be exactly at the critical point. So I don’t have time to explain the details, but this is something that, sort of, intrigues us.

Let me take one moment to have a little fun with this. I told you that something surprising was happening, right? I have a network of neurons that talk to each other in very complicated ways. Every neuron increases the retina—it’s not as bad as the cortex—but it’s still true that individual neurons are connected to many other neurons, and many neurons can receive common input from the same neuron and so on, so it would seem like there’s all sorts of common notorial possibilities, and yet, I’m telling you, that if I just keep track of pairs, everything works out. In physics, this is a very common idea; you could argue that almost all of the complexity that you see around you ultimately arises from paralyzed interactions. There are electrons, there are protons, right? And the force between them, and you get kind of stuck and you keep going, right? So it shouldn’t surprise you that paralyzed interactions can do a lot, but at this phenomenological level, you can think of many reasons why it wouldn’t work. So one of the place tags Greg Stevens and I tried to lay around with a little bit, you’ll notice the in-progress and the dates don’t match, but that’s actually correct because we haven’t made much progress recently, but, so we decided to play around and look at four-letter words. So you go through the chosen, of course at-random of all possible four-letter words. You go through one year of writers and you restrict your vocabulary a little bit so you can sample well enough, and you think about the four letters as being like four neurons, except of course instead of just being able to spike or not spike, they have twenty-six possibilities, so I have a sort of network of letters that forms a word. So I remind you that if you actually tried to write out the probability distribution that has twenty-six to the fourth elements, which is half a million, but if you make this model that only keeps track of paralyzed interactions, well, there’s twenty-six squared pairs, but then there’s six ways of choosing pairs out of four letters, so that’s four thousand. Still a lot of parameters, but a hundred times fewer than it could be. So, this one you all know; if all the letters were used equally, then there would be about eighteen-, nineteen bits worth of entropy to a set of words. There aren’t two-to-the-eighteenth four-letter words, so you know that’s not right. First of all, you know that not all the letters are used equally, so that cuts you down to here. There still aren’t that many four letter words, so you know that the distribution is not uniform over all the possibilities; there are many people in the audience who know much more about this than I do. So let’s compute the actually entropy, and we can discuss how well we estimate it, and well, it’s about half. What that means is that there’s a gap between the entropy if all the letters were independent and the entropy that you actually see, and that gap, which is about seven bits, is how correlated the letters are. When you’re building a model that keeps track of the paralyzed correlations, but otherwise is as random as possible, what you’re doing is computing the maximum possible value of the entropy that is consistent with the paralyzed correlation. So you can think about that every time you add more correlations, you always push the entropy down. So what’s surprising is that, although spelling rules, especially in English, as seen by non-native speakers, seem to be horrible, -they’re very combinatorial, you know, this here, that there, sounds terrible- but actually, by keeping track of paralyzed interactions, we squeeze out 87% of the entropy that’s there. So this suggests that maybe these tools are more powerful than we sometimes give
them credit for. By the way, inevitably you will end up assigning non-zero probability to words you didn’t see in the data base, but that’s not so bad, because remember, the data are just the most common words. So what’s interesting is to look up the nominally non-words that you generate, and what you discover is that some of them are actually words, or at least slang, and some of them- they aren’t words….but maybe they should be.

Anyhow, this is still at a stage where we’re playing around, but it’s maybe to leave you a little intuition on how it’s going on; how it’s possible for these models to work so well. So let me try to wrap up here.

I tried to follow two very different tracks. One is to ask “What problem is it that the brain’s actually solving?” , and the other, which I was actually talking about, is “How do neurons cooperate in networks?” And in some sense, this is a kind of top-down view of neural computation, and this is, well if not bottom up, than at least lower-down up, right? You’re trying to take neurons and build up something as opposed to some abstract computation and work your way down. So I gave you some examples of the kind of computations the nervous system has to do: classification or future-estimation. And, in these examples, it seems as if the system comes close to some well-define notion of optimal performance, and there are many other examples along this line, and it connects this whole line of reasoning- connects with other lines of reasoning and thinking about neural computation and perception and so on; I didn’t have time to draw all the connections for you. But what’s important that if you take this seriously; if you take this observation of near-optimality, then you have a theory, automatically have a theory of what the brain should compute in order to solve these problems, and then you could ask whether the system actually computes that. And the key qualitative prediction that comes out of this is that what you should compute almost always depends on the context in which you are working. I didn’t emphasize that, but it’s a very important part of it, and that actually is one of the experimental handles you have, and also one that which I think ties even simple problem, low level problems, to much higher level questions, because if you want to solve a low-level problem optimally, you actually have to have a handle at least on the context in which you’re solving it, where context in this case usually means something about the statistics about the environment that you’re working in. And so, you can’t really manage optimal estimation unless you’ve done a good job of the learning the aspects of the underlying probability distribution. So not only are the problems that get solved at a low level, or perhaps in simple organisms are analogous to the problems that we solve on a more cognitive level, but perhaps imbedded in these problems is a problem that surely a more cognitive problem- mainly that of learning the underlying statistics. But I would claim tat this line of reasoning has a very serious flaw, and that’s the laundry list problem. I tell you that the nervous system, let’s say, extracts motion, or it allows you to estimate the location of a sound source based on the signal from your two ears, or it computes where you’re supposed to put your hand, what the motor program is, for reaching for something, or whatever, and in each case I tell you there’s a motion of optimization. But why these features? Why are these the things that you compute optimally and not anything else? And so, what I think the observation that the nervous system is capable of solving problems with a performance that’s close to optimal, really sharpens a much harder questions, which is “what is it about the things that the brain is good at doing that somehow ties all these things together?” And the tentative idea is that
almost all the things that are entirely useful have something to do with being able to make predictions. And this is sort of at the edge of what we’re trying to do now. On this side, essentially the observation is that even if correlations between pairs of neurons are weak, there are lots of pairs, and what we learn from simple models of statistical mechanics, is that if n-elements are all talking to each other, then what you mean by “weak” is not “weak” compared to a coefficient correlation of one, it would be “weak” compared to a coefficient correlation of one-over-n. And actually, if n gets big, it’s easy for the coefficient correlation to be larger than one over n. In fact, it’s easy for your threshold of significance in a finite experiment to larger than one over n, which would mean you could completely miss correlations which are the evidence for a very dramatic collective behavior of a system as a whole; you would declare them to be too weak. So there’s something about this line of reasoning that I really like, which is that all these exotic things, about analogies between neural networks and spin-glasses and stored memories and ground-states and all this stuff that seem like idle theorist chitchat, because experimentalists kept saying “I don’t see anything like this”. Well actually, they did. Because those patterns of weak but widespread correlations are exactly what these models produce. In some sense, the minimal structure that is implied by those correlations is this every exotics theorist playground; it could be richer than that. And that I find really intriguing. Especially intriguing as you think about “we could do this on the retina”- imagine if we could do this on the cortex. And of course, the real problem here is that, the top-down

[lecture tape picks up again here]

and bottom up approaches eventually have to meet, and the problem I would note is that, over here, we’re seeing that there’s certain types of problems that you’re very good at solving, and in some sense the system is poised at a very special point, where it manages to achieve computations which are as reliable as possible given... Over here, what we’re finding is that the networks that we’re able to analyze also seem to be poised at a very special point, and it’s not impossible that these points are related to each other. And what I hope to have done, in only slightly more than the usually allotted time, is to have illustrated for you that there are problems in neural information processing, which are clearly problems like the problems of cognitive science; They involve classifying things, estimating things, and they involve understanding how large numbers of neurons cooperate in order to produce something together that no single one of them can do on its own. What we’ve tried to do is find instances of these problems which are accessible; we don’t know how to go into the cortex and do the analysis we did on the retina. On the other hand, I think that the mathematical ideas that have sort of floated in the literature get a very concrete instantiation as we try to look up what is admittedly a simpler problem, and that these can provide us with a guide for thinking about the richer problems, and of course, I can’t claim how reliable that guide is going to be, but I hope that, if nothing else, what we succeed in doing is sharpening the questions that we ask, even if the answers turn out to be different.
Questions and Answers

Question: The observation is that these models have the property that there are these ground states, which, if you think about the probability, are basically local peaks in the probability. So, you might ask “how many of them are there? How peak-y is it? And how do you get from one to the other?”

Bialek: So the last one I’m just really going to punt on; we really don’t know much about the dynamics yet, so I don’t have anything to say. “How peak-y are they, and how many are there?”- something that really surprised us, although whether or not we should have been surprised is another story. The experiment involves recording from forty neurons, and then you find that there are sort of four or five non-trivial states, out of forty. So it’s an eighth, which, by the way, isn’t so far from the capacity of the Hopfield model but never mind, what’s intriguing that when you go to, if you try to extrapolate to a much larger network of one hundred and twenty neurons, now there’s really a proliferation of states, and in fact, if I ask you about the entropy of the whole distribution, one third of that entropy is really contained in what basin you’re in. So what both intrigues and worries us is that, if you move to very far beyond the current generation of experiments but well within the range of the next generation of experiments, the richness of this picture expands enormously, so we’re really excited because this will get tested. It’s also important because, I said, sort of, all cells are correlated with all other cells, that’s true, obviously, within some radius. We’re in the retina, right? If a cell’s looking over there and over there it’s not going to be very correlated. So that radius, within which the correlations are spatially homogeneous, before you start to see the fall off, contains an order of two-hundred cells. So we’re intrigued by the possibility that the combination of the strength of correlations and the numbers of cells are such that by the time you get to two-hundred, things will be really interesting. And, it is feasible to record from all two-hundred of those cells: it will take a year.

Question: …in the very beginning of the talk, you gave some very nice examples where the information available in the sensory input was used in and by the organism detection…implies… this time around. There are also examples, for example, of neurons in primate cortex crossing over at rates far above… the information clearly sends who transmitted the cortex and loss of behavior. And I wonder if you thought about those kinds of examples, like the…….

Bialek: So, to the extent that what you’re interested in is anything other than a perfect reconstruction of anything you ever saw, then you throw the information away. For instance, I think in the example you give that it’s clear that you have to preserve visual information with very high time resolution, until you’ve sorted out questions of what’s moving and what’s solid. So it’s possible that I could make the same statement about the retina, right, but there’s very high time resolution information available there which is lot behavior, but that’s not quite fair, because were you to corrupt the temporal precision of each channel, then subsequent computations that involved comparing those channels to decide whether you’re actually seeing the same thing or not would probably fail. So I don’t actually know whether that’s- it’s hard- so what’s weird about the examples that I
choose is, I think, what allows us to select them, I hope is not just a bias on our part. It’s rather that, these are situations that for one reason or another, it’s pretty clear what’s interesting and what’s not interesting. And so, I can tell you that the information about the interesting thing is perfectly preserved, or I can compare that to some meaningful limit. I mean, look, if I look at am motion sensitive neuron in a fly’s visual system, this is a wide field motion sensitive neuron, it’s thrown away everything about spatial structure, so of course things are being thrown away, and I don’t know whether or not that by itself constitutes an example of sub-optimality or not. Tom.

Tom: So, it’s unclear to me, what you say about neuro-transmission, whether I’m supposed to be thinking about a stem-glass or something else. Is it necessary to get the sort of behavior you have anti-correlations……

Bialek: So, an empirical statement is that not all the correlations are positive. It’s a little hard to see here…this was not made for really scientific purposes; there are weak anti-correlations. In fact, if you look in the model you end up building, forty percent of the triangles are frustrated. So the question was, whether these kinds of models have many different behaviors depending on the signs of the interactions. So for instance, if everybody interacts with everybody else in a positive way, then you can have a transition, but the transition is to a simple ordered state, in which everybody’s on or everybody’s off. Here you have competing interactions, and the reason you know that’s going to happen is because some of the interactions are positive and some of them are negative. And in fact, if you sort of think about any three neurons, there’s about a forty percent chance that the interactions will be frustrated, that is to say, these two guys will have a positive interaction, these two guys will have a positive interaction, and these two guys will have a negative interaction, so they can’t all decide exactly what to do, and that’s ultimately what drives the emergence of these multiple states. And Tom’s question is “to what extent is that essential?”, it’s certainly essential to some of the behavior, we wouldn’t get the multiple ground states. Whether you could imagine a system that was poised neuro-transition, but the transition was not into one of these complicated spin-glass states, but rather into something more conventional. I don’t know, find a network, and see if we can…..

So we’ve been, up to this point, embarrassingly empirical about all this, right? I mean, there’s obviously a whole set of theoretical questions, well, you know, we’ve written down all these models, let’s say more about they’re properties, let’s be professionals about it. But our focus has tended to be on “Are we sure that these models are actually describing the data.”

::Garbled question::

Bialek: Well, it’s more, close to being a mean field-spin less than to being sort of ::garbled::. Sorry, that one doesn’t need the translation to me.
Bialek: Right, what you would expect, is that if the new movie were of statistically the same structure as the old movie, not exactly the same movie, but in some sense drawn out of the same ensemble, then to the extent that the correlations among the neurons come out the same, and the mean firing rates come out the same, you’d predict the basins were the same, and it’s an interesting question, whether you’d visit them again.

Bialek: So, there’s an interesting—there’s a larger question behind the one you’re asking, which is that, in addition to the fact that networks of n-neurons have two to the n states and experiments tend not to be of length to the end, it’s also true that networks of n-neurons typically, at best, you sample m less than n neurons, so how does this whole picture fit...forget whether they get killed or not, suppose you just don’t see them? We already have a problem. To what extent should this description work, even when you only see part of the network, which we surely are doing. What you can do theoretically is say, let’s start the model for forty neurons, which is all we have, and then let’s start numerically taking out the neurons and averaging over their behavior, and then asking whether or not the model we end up with is what you might have thought you’d get if you’d pretended that the twenty neurons left were the only ones that we’re recording from when you tried to reconstruct this model. There’s a number of reasons why you might think this wouldn’t work; essentially, you have three neurons you’re recoding from, and one neuron that you’re not, and as this neuron does it’s thing, it can basically tie all three of these guys together, and that gives you an effective interaction between three neurons and we’re only keeping track of pairs. But it seems like the parameter regime that these models are in is one in which that’s not a very big effect, and so your guess about how strongly these two neurons are interacting with each other is, of course affected by whether or not you’ve seen all the neurons or not, right, so in some sense the fact that this cell can talk to this one who talks to this one, if you don’t see this one, it means that these two guys are interacting a little more strongly than you thought, right, because their interactions are mediated by the other one. So those kind of effects do happen, that is to say, your ideas about what the interactions are depend on whether you’ve seen all the cells or not, although I must say that it’s not completely obvious why that should be true. I should not that in this system, where we’re looking at the response to fairly natural movies, trying to interpret the interactions that we get as being, sort of, physical connections; this is a very bad idea, ok, for a variety of reasons. One I just mentioned, what if there’s a neurons you’re not recording from. There are experiments that have been done in the primate retina, by John Schlanz and E.G.H Lanisky at SAW, where they’d looked that the correlations in the dark, with everyone just sitting there, and they’ve used similar mathematical methods and find again that these sort of maximum entry models do a very good job of describing things, but what’s interesting that in that limit, it seems that if you pretend that the neurons only interact with their nearest neighbors, you actually do a pretty good job. And actually, that corresponds reasonably
well with some of the anatomical features of the class of neurons that they were talking about, so it’s possible that if you take away all the stimuli and let the system just be tickled by all it’s background noise, you then get things which are perhaps less interesting in terms of network dynamics, but more easily interpretable in terms of anatomy. And then, some of those questions about “what does this connection mean, and does it change when I only see part of the system” and everything, become much sharper.

:::garbled question:::

**Bialek:** The question is about the model for four letter words, so let me answer the question by sort of contrast to the traditional approach, right? So a standard approach is, I’m going to ask what letter this is, and I say that the probability of using letters at this point depends on the k-previous letters, ok, so it kind of k- for the mark-off. So, in that view, it could depend on any aspect of your history right? So if you want to have a memory of length-k, you need twenty-six to the k parameters. So what we’ve done here is to say, your memory as it were can reach across the whole word, so you can have a memory of one, two, or three, but you don’t need twenty-six cubed parameters for that, right? Because they’re only allowed to interact in pairs, and in some appropriate sense, the interactions from the pairs have to add up. So, in this view, if you want a memory say in the conventional view of the simplest implementation of the k-per a markup model for letters, if you want a memory of length-k, you need all the combinatorial possibilities of length k, and so you have twenty-six to the k parameters. Here, since we’re only allowed to paralyze things, you can go and have a memory of length k for a cost that is linear in k; you get k times twenty six squared parameter.

:::garbled question:::

**Bialek:** Right, but what I’m saying is it’s not obvious….what I’m telling you….so your observation is correct; a big part of the restriction of the entropy is simply that most things don’t make words- forget about the probabilities, it’s just that most combinations aren’t real words. So you could say that all the entropy reduction is putting zeros in the right place during entropy distribution. What I’m telling you is that I can get the zeros in the right place by just keeping track of pairs of letters, which means that it’s not actually combinatorial, or at least some of it’s not combinatorial.

:::garbled:::

**Bialek:** So, I gave you this as an example partly because I figured that in a cognitive science audience, there are probably more people who know that language and maybe it would be fun. I have absolutely no idea whether this is actually a good way of describing language or not. The point I’m trying to make is that this restriction to paralyzed interactions really provides with a cut through the space of models which is very different from some of the standard ones in this context. Whether it’s really a productive cut or not- I mean, checking four letter words is really not a way to tell - it’s just by way of illustration. In the case of neurons, I think we’re on the road to show that it really is a productive cut through the space of models.
We’ve come to the end of our time.