Brain and Mind
May 13, 2004

Gerald D. Fischbach, MD
Neuroscience and Neuropathology—Converging Streams

Introduction by Lee Bollinger

President Lee Bollinger: This symposium is the result of the very, very hard work of Professor Tom Jessell and Dr. Joanna Rubinstein, and I want to thank and acknowledge them, and would like to thank all of you for coming. This is a great testament to the general perceived importance of the subjects of this symposium. I want to thank all the speakers who have come to participate.

I want to take this occasion just to announce that Columbia will be launching—we are launching, as of this moment—an institute for neuroscience that will be part (eventually) of the major center for the study of the brain and behavior. We all, I think, recognize in the academy the extraordinary advances that have come just in the past few decades, the past decade in particular, because of the discoveries around the genetic code. Where that will take us of course we don't know, and we're making very, very significant investments across the country in trying to advance knowledge as a result of that new knowledge. But the study of the brain and how it works is clearly central not only to the curing of disease, but also to the understandings that we bring to every, virtually every, area of life: social sciences, the professions, and the humanities. And it is Columbia's goal to try to bring as many scientific advances as we possibly can to this area, and also to integrate it with other areas of knowledge.

Dr. Tom Jessell will be our leader in this new venture, and Dr. Gerry Fischbach, executive vice president for health sciences, whom I will introduce in just a second, has, of course, himself as a neuroscientist been trying to bring this about. We've been working together on this, so from the health sciences to the University this will be a major effort on our part. We will make it one of the central goals of a capital campaign that will open sometime in the next year or two.

I would like to now introduce Gerry Fischbach, who will take over the next panel, or the first panel. Gerry, many of you know, I know he is well known to all of you who work in the area of neuroscience and in health sciences, and before joining us at Columbia in 2001, he was the director of the National Institute of Neurological Disorders and Stroke at the NIH. He was chair of the Department of Anatomy and Neurobiology at Washington University, and also served as chair of the Neurobiology Department at Harvard and Mass General. Gerry is to his core an academic, an intellectual, a scientist, a very, very caring human being who...
believes deeply in the power of knowledge and the pursuit of knowledge, and has
taken on this role of heading up the medical center. Unfortunately, every now and
then, financial matters intrude on what is really the source of his dreams for the
institution, in terms of the academic vitality. And we do our best here to keep the
focus really on the academic side, and it's wonderful to have someone like Gerry,
whose life is built around that goal. It is my pleasure now to give you Dr. Gerry
Fischbach.

Welcome and Introduction

Gerald Fischbach: Thank you very, very much, Lee. Financial matters are minor
compared to the study of the brain and behavior, so that's what we're going to
focus on. And I believe you're in for an extraordinary treat today because Tom and
Joanna have really assembled a wonderful group who will illustrate for you the
spectrum of neuroscience and the reason it is a proper function of a whole
university. What you will hear today, this morning, this afternoon, and tomorrow, is
a range of subjects of how we're approaching modern brain science, ranging from
studies at the atomic level, reaching out toward chemistry and nanoscience about
the structure of key molecules in the brain, through an understanding of how
groups of nerve cells work together to form small circuits in the brain that we know
now account for simple behaviors. How they're wired together during development,
and how we are beginning to understand more complex functions, like sensation
and perception, the thread of life that keeps us whole, and a sense of awareness
that is the function of memory. And later on today we'll hear more about
developmental disorders from a molecular point of view, and also using tools of
imaging to understand how devastating developmental disorders of children come
about. And here the discovery of a gene is really a blessing, it gives you a clue to
mechanisms and how one might approach the disease. And later on you will hear
studies of one of the parameters going up the scale of integration, what really are
the parameters that one must understand to really analyze mental life and mental
phenomena.

My role here is to thank you for coming—there are 2,000 people registered—to
add to my thanks to Joanna and to Tom, to thank a few people in the audience
who have contributed already to our neuroscience community, the Kavli
Foundation with its emphasis on neurocircuits, the Taub Foundation with its
emphasis on neural degeneration; and, I do want to add that Richard Mayeux of
the Taub Foundation will play a key role in the evolving Neuroscience Institute
along with Tom, as mentioned by Lee. And there are many others who have given
us an enormous boost and the courage to create a university-wide institute.
A Primer on the Nervous System

My second role here is to give you a primer on the nervous system, and I'm going to do that by discussing two disorders: myasthenia gravis and Parkinson's disease. And my hope here is to show you the interplay between fundamental science—some people would call this curiosity-driven science—and the real approach to human disorders, which is becoming more and more common these days. Certainly my experience at the NIH said that we really are on the verge of translating many fundamental discoveries into useful therapies, and I'll tell you a little bit more about that. Hopefully I can do all this within 15 minutes.

Now I would like to have the first slide. Can I have the first slide? I can see it.

This is the object of our study for the next day and a half. It's certainly one of the most complex and most mysterious objects in the universe. It's not complex because of its size, it's only about 3 pounds. You may know that Albert Einstein's brain was 2.75 pounds. Mine is just a little bit bigger than that. But it is complex because of its component parts. There are hundreds of billions of nerve cells in this brain, and there are hundreds of trillions of connections between them. But the really important part is that these connections between the cells are far more complex than their number would imply because they change with experience. Not only are they laid down by genetic programs but they're influenced by everyday life, by what you are listening to right now.

Now can I have the overhead? I want to simplify for a minute by taking a section through this brain and looking at a cross-section through the brain. I want to tell you where these two groups of nerve cells that are involved in myasthenia gravis and Parkinson's disease reside. Coming down from the cerebral hemispheres there are groups of cells in the brain stem, which continue onto the spinal cord. One group of cells that I will talk about use a neurotransmitter called dopamine, and they reside high in the brain in a region called the midbrain. The other group of cells I'm going to talk about are motor neurons that reside in the spinal cord, all up and down the spinal cord, that send axons out from the spinal cord to the periphery to innervate muscle, in some cases glands. These are the effectors, the output of the brain, they synthesize all the information from above and send processes out. Now these processes, called axons, are long. They're thin compared to the cell body from which they come, but they are extraordinarily long. If drawn to this scale as you're seeing on the board, an axon in the neck that goes out to a tip of your finger would extend up to Washington Heights and back on that scale.

So the way the nervous system communicates is it sends from its cell body brief electrical pulses called action potentials that are conducted at a constant velocity out the axon, and they travel that long distance in a matter of a hundred milliseconds, a tenth of a second. And in the case of muscle, they arrive and form very unique connections on the muscle, and it's this connection that I want to dwell on for a minute.
These connections are called synapses, and the synapse on muscle is probably the best understood synapse in captivity. I could go back to the slides again. It's in the periphery, it's accessible, and we know more about its neurotransmitter, acetylcholine, than we do about any other transmitter in the brain. Indeed we know more about the detail, the structure of the synapse, this particular synapse, than any other. When the impulse arrives at the end of the axon after that long trip it does not continue directly across and stimulate the next cell in line. There's not enough electricity in that impulse to do it. Instead, when it arrives at the end of the axon, it causes a chemical reaction and the release of small molecules called neurotransmitters into the synaptic cleft, into the gap between the nerve terminal and the next cell in line. It is this release of a chemical that leads to enormous amplification of the signal, and in the case of the nerve-muscle synapse, to assure transmission. Every impulse that arrives at that end of the terminal leads to an action potential and a contraction in the muscle fiber.

Now modern technology has allowed us to study these molecules in the muscle, in the postsynaptic membrane called receptors for the neurotransmitter. One can detect the action of a single molecule of transmitter by activating a receptor. And you can see in this physiological trace, I hope, that a receptor is either closed or, when it binds neurotransmitters, opens briefly and closes again. If you're close enough to see the scale here, these are pulses of current that measure in the order of picoamps, one-billionth of an ampere that drives your amplifiers at home. And we can now detect that. And it is possible to look at these receptors at the electron microscope level and get an idea that they really are channels looked at face down, and in this figure to the right an image, a ghostlike image, of what a receptor might look like in real life. These receptors bind transmitter molecules and they open and close briefly, allowing these small pulses of current to occur. And when these pulses all occur in synchronise there's a huge current that enters the cell and excites it, in this case the muscle to contract and causes movement.

**Ion Channels and Myasthenia Gravis**

Now you'll hear in the next lecture a much higher resolution study of these transmembrane ion channels from Rod MacKinnon, but for now I'm going to move on and just describe to you how these channels are tremendously important in certain diseases. This synapse is a sure thing, but at a disease junction in a disease called myasthenia gravis, the number of receptors is decreased, and I want to illustrate for you how science, this type of science, has shed enormous light on the disease myasthenia gravis.

It was realized early on that one could purify these receptors away from the membrane environment. That was accomplished by quite independent studies of the snake called *Bungaris multicinctus*, also known as the banded krait—a real public health problem in Taiwan—and in an effort to get rid of this snake and the problems the snake was causing, two workers in 1966 purified from its salivary
gland a protein molecule that acted as a toxin. It looks like the snake itself on the right. And it was found that this toxin bound very tightly and very specifically to the type of receptors that are in the membrane of muscle cells that can bind acetylcholine. And with the use of this toxin labeled with radioactivity it was possible for the first time to quantitate numbers of receptors. And in 1973 workers at Hopkins used the toxin, and here you can visualize the toxin by these grains in an autoradiograph, dark spots where the toxin illustrates the accumulates of acetylcholine receptors in a normal muscle and a myasthenic muscle. And I hope you can see in a glance that the number of receptors is drastically reduced at the junction, the first real, hard evidence that there is diseases of the postsynaptic membrane. And in this case the decrease of receptors leads to a profound weakness, beginning in the facial muscles, droopy muscles, but can become extremely severe if it involves muscles of swallowing or muscles of respiration. Twenty-five thousand people in this country each year are diagnosed with myasthenia gravis.

And then a mechanism was suggested, again through disinterested curiosity-driven studies. Workers at the Salk Institute wanted to raise antibodies in rabbits against this protein so they could further study the structure of the acetylcholine receptor. And when they injected a healthy-looking rabbit like this and waited a month or so for the antibodies to accumulate and went down and examined the rabbit in the animal room they found that the animal was nearly paralyzed, ears droopy, floppy, and until that rabbit was treated to preserve acetylcholine it could not stand. And this gave the investigators the idea that the decrease in receptors at the postsynaptic membrane was due to an antibody attack. It’s now one of the best understood immune disorders, autoimmune disorders, in the literature.

Now, in turn, our understanding of myasthenia has stimulated enormous amounts of science. We now know more than twenty different diseases, congenital myasthenias, known as channelopathies, where muscle weakness occurs because of defects in the postsynaptic membrane and the presynaptic membrane. This stimulated my own interest and some of my own research, interested in the fact that these receptors are labile, they can turn over. We were interested in how receptors accumulate during development and how they are maintained in the muscle membrane. We learned how to grow nerve and muscle cells in tissue culture, and we purified—I'm summarizing ten years of work quickly—we purified a protein shown here diagrammatically known as neuregulin, which when added to muscle fibers (shown on the left in this autoradiograph, again the black dots signify acetylcholine receptors), caused these muscle cells to synthesize and accumulate many more acetylcholine receptors. And we believe this is a vital factor made in motor neurons that maintains the integrity of the muscle membrane, the postsynaptic membrane.
New Therapy for Parkinson's Disease

Now I will end by talking a bit about Parkinson's disease, and I want to show you a very moving sequence of a new therapy in this disorder, again based on years (about thirty years), of study in nonhuman primates. It promises to be a really extraordinary if not miraculous cure.

You remember these were a different type of nerve cell high up in the brain, in the midbrain, that doesn't use acetylcholine but uses a neurotransmitter called dopamine.

"By the time we met Sybil this past spring even a simple task like eating breakfast was a frustrating battle with her own body. How can you live your life when you're shaking so much? 'It's extremely difficult, sometimes even overwhelming. And I get very emotional, sometimes I cry. It's just a hard thing to do.'"

"The disease had gotten so bad that sometimes her muscles froze completely making her face almost expressionless and her legs almost useless. The woman who once was always on the go could barely move, confined to a wheelchair. But as the disease got worse the drugs did less. 'One of the things that this disease has really accomplished is it has really robbed me of my . . . my dignity, my self-esteem, because it has taken away my independence. When you have to be totally dependent on other people to do everything for you, you actually lose your self-respect, you feel worthless, and that's what Parkinson's has done for me.'"

"But on the morning of the surgery Sybil was confident, even as nurses fitted her with a head frame to help guide surgeons during the operation. Deep brain stimulation comes down to one thing, location. Finding the target area holds the key to Sybil's future. The device electronically records the activity of individual cells within her brain so the doctors can actually hear the trouble spots. That noise is the sound of a neuron misfiring. 'Can you move your ankle up and down? It's hard to do for her.'"

"Finally the doctors think they have placed the electrode in the exact location that will help Sybil. When they turn on the stimulator she suddenly shows an astonishing range of motion in her legs and hands. 'Can you pick your leg up at all? There you go. Up and down? Drop it down now, now pick it up. Great. She couldn't do that before. How about the hand? Can you open and close your hand for me now? Good.'"

"She was doing so well that Dr. Vitek had a surprising suggestion, more surgery. He thought that a second implant in the right side of Sybil's brain could get rid of the tremor in her left hand and might improve her stride. 'The minute they got in and they hit the target, it's like magic. "Okay, open and close your hand." He said to me, "Open your hand and close your hand," so I knew right away. It's amazing.'"
"Sybil now had two stimulators, one for each side of her body. "How do you feel?" "Great." "What can you do now that you couldn't do before?" "Oh, good heavens. I feel as if I've . . . I've been given my life back. And I'm taking hold of it.""

Now this is an extraordinary accomplishment. It's not a procedure for everyone diagnosed with the disease, but Sybil Guthrie, who I had the real privilege of presenting to a congressional subcommittee in Washington, has had these electrodes in place for more than 12 years. And there are studies throughout the country based on this to understand the mechanism and the viability as a general therapy. Who's going to pay for these procedures is quite another story, but it indicates that despite the degeneration and loss of cells, a deep knowledge of brain circuits and where to stimulate may be able to overcome and correct these defects.

I want to end by saying that we really are at the beginning. This will be a challenge for all great universities, and here's a reason why: This is a picture, a diagram, of the electrodes placed in Sybil's brain. You can stimulate from any one of those four dark spots. When the electrodes were switched in a patient (in a study performed in France and published in the New England Journal of Medicine) this following sequence occurred: These four images occurred within a matter of minutes of each other. When the electrode position was switched within four minutes, excuse me, within a matter of seconds this very alert patient became pensive, and within a few minutes after that became tearful, suicidal in fact, uttering phrases such as she could not go on, this disease had become overwhelming, not due to the sensation of the stimulation, but simply due to her emotional tone. And then when the stimulus was stopped, within a few minutes she recovered and became positively euphoric, illustrating the profound influence these dopamine neurons may have on emotion and mood in addition to movement. It also illustrates how little we know about the brain circuits in this area, how much we have to learn. And it raises profound ethical issues, which I can tell you will become more and more important as our techniques advance, in understanding what really is involved in informed consent and being ever alert to adverse events during the course of experimentation on the brain.

Question and Answer

Gerald Fischbach: I'm going to stop here and maybe take one question or two, and then move on. Thank you.

We thought despite this big audience it would be useful if there was a question or two after each talk, and if not we're going to move on—Yes?

Man: Is it possible to tune the stimulus, can you . . . instead of having all or none, is it possible to tune it?
Gerald Fischbach: Yes, in fact one of the great challenges here—that movie came from Emory University, but it’s a technique that’s being . . . yes, the question is: Can you tune the stimulus? And I’ve said that this technique—the film you saw was from Emory University—originated in France, now is being practiced throughout the country, including at Columbia in our neurosurgery group. And the art is, after the electrodes are placed, to tune the stimulus, both the frequency and the magnitude of the stimulus, and that’s something that can be done in the doctor’s office after the electrodes are placed. And there are many surprises in store I think when that tuning process occurred.
Introduction by Gerald Fischbach

Gerald Fischbach: I would like to move on and introduce to you Rod MacKinnon, our next speaker. Rod is a friend of many people at Columbia. He now is at Rockefeller University, and someone I've known for many years. Rod MacKinnon began his life as a physician, his scientific life as a physician in general medicine, and became interested in biophysics—a renewed interest, which he had as a student, and was among the most profound scientists studying the function of ion channels, these proteins in the membrane, one example of which I gave to you. But after years of study and quantitative analysis of these . . . the function of the ion channels, Rod decided that he wanted to hold a channel in his hand, and he decided to become involved in the structural biology, the chemistry and structural biology, of ion channels. And in a matter of a few years, a very short time, he accomplished this enormous feat of isolating, crystallizing, and studying the structure of an ion channel, an accomplishment that has won him, I believe, universal acclaim. Certainly if you look in your programs you will see the awards he has won, from the Horwitz Prize at Columbia to the Lasker Award, the Gairdner Prize, and last year, the Nobel Prize in Chemistry. So Rod would you come and give your talk?

Ion Channels Generate Electrical Impulses

Roderick MacKinnon: Okay. Thank you very much. It's a real pleasure to be here to tell you about potassium channels.

So I start with this picture, set of pictures, entitled "Ion Channels Are Nature's Electrical Impulse Generators." And here you recognize a brain and here a cell, a neuron, from the brain. And this is an electrical impulse. And Gerry Fischbach mentioned those electrical impulses to you. And what generates those are molecules called ion channels, and there's a picture of one here. And these are small, very small, atomic-scale molecular machines, and they make the electrical activity of the nervous system.

Now let me just begin with a little bit of summary: All cells, including neurons, have a membrane surrounding them, and in that membrane there are two classes of what we call ion-transport proteins. One of those is pumps, shown in a cartoon version here on the left, and the other class is called channels. Now what the
pumps do is they use energy, typically chemical energy, for example shown here, hydrolysis of ATP, to pump ions against their electrochemical gradient. And so what these accomplish is to establish a gradient across the membrane, and it's depicted here showing on the outside of the cell, red ions, and on the inside, green ions, and this pump spends this energy to separate the red ions, which would be sodium on the outside, and the green ions, potassium, on the inside.

Now why does this happen? Well this separation is in a sense a form of energy, usable energy, because imagine if the ions were given the chance to run back down, they'd run down their concentration gradient. The reds would run back in and the greens would run back out. And that's what the channels do, that's where they come in. So here we have a channel that makes a pathway, and this channel is a pathway specifically for green ions, potassium. And what happens when this channel opens, potassium runs out, wants to run out, and because the potassium ion carries a charge, when it starts to come out it leaves a negative behind and carries a positive out, and so it polarizes the membrane electrically. It's like an electrical capacitor. And so you get a charge across the membrane. And that . . . there are other things added to this, but in the simplest sense, this is where the electrical impulse comes from; it's a polarization of the cell membrane. And Gerry Fischbach told you about neurons that originated in the brain that ran down the spinal cord, they made contact with motor neurons in the spinal cord, which sent axons out to the muscles in the perimeter, and those axons carry electrical impulses that correspond to movements of this kind of polarization. That's what the electrical impulse is.

**Structure of Ion Channels**

Now what I want to address for the next little while is: What does an ion channel look like, how does this work? When you look at this cartoon, I showed you a pathway across the membrane and I said green ions go through to carry a charge across to polarize the membrane. And if you think about that you realize that this channel, although in a sense it's a very simple thing because it's letting the ions just run downhill so to speak, green is concentrated, wants to run out, this channel when it's open can only let the green one through. If it let the red one through, this system of polarizing the membrane would fail. So we can see that the channel has to . . . you start to get a feel, but the channel has to have special properties, and one is that it has to be selective, only a specific kind of ion can cross. And another is to make a proper useful nerve impulse, it has to conduct the ions very fast when it opens.

And so how does it work, how did nature make these little machines? And one way we can start to figure out that is to see it, because we all like to see things, and when we look at things we can get a feel for how they might work. So that's what this is about. I want to show you what some of these machines look like.
This is the pore of a potassium channel. This potassium channel happens to be from a bacterium, not from a nervous system. But what's interesting is that across the tree of life many molecules that are used in various circumstances are highly conserved, and it turns out that many of the ion channels, in fact most of the ion channels, are highly conserved across the tree of life. I'll come back to this in a little while, near the end. But when we look at this we can learn, in fact, about potassium channels that are in our own nervous system by looking at the structure of a potassium channel from a bacterium.

Now what you're looking at here are four subunits in different colors. And I've got this thing rotating about a central axis that is the pore where the ions run. Outside the cell would be here, inside the cell would be down here, so the membrane would run from here to here. This is very small—a very pretty big picture—but in fact the distance here is only forty angstroms. It's extremely small—much, much smaller than a single cell. It is spanning the skin or membrane that surrounds a cell, which if you look at a cell like an onion the skin around the cell is like a peel of an onion, so that will tell you how small this is.

Now how does it work? Here's a simpler view of it: If we take one subunit away from the front and from the back so that we only see two, this lets us look down the middle. And what you see is a pathway running from the outside to the inside, and there's a particular structure here. It is, in fact, to a structural biologist . . . you know, it tells us some things about how it works. It's a kind of a . . . I think looking at these is very interesting because, in a sense, nature, I think, is very beautiful on many levels. We can, you know, look at sunsets and mountain ranges and they're very beautiful, and you look at organisms and they can have very beautiful shapes and forms, and even down to the atomic level you see nature has beautiful shape and form. And I think this kind of molecule does. The pathway runs down the middle, the blue is actually ions in water along the pathway. And so by looking closely at this we can figure out what nature is doing with this little device that runs ions across the membrane to make electrical impulses.

Now remember, one feature of a channel that I showed you in the beginning is it has to be selective. So why does this channel only let a potassium through? So if we look closely at this region right here, it's a little segment of the ion conduction pathway, we can see that it has a particular structure. For now . . . for its fine details, I won't go into now, other than to say that the green positions are where potassium ions bind when they hop from one site to the next as they cross this thing called the selectivity filter of the potassium channel. And these red sticks are oxygens from the protein, and they reach in and they hold a potassium ion, so in a sense this is a register of four positions where ions can hop from one to the next as they cross the membrane.

And I mentioned that these molecules are very conserved across the tree of life, and I want to show you how conserved they are. Notice that this sequence I have labeled as T-V-G-Y-G, that's the single amino acid code for types of amino acid,
threonine, valine, glycine, tyrosine, glycine. And if we look through the gene
databank at potassium channels and ask what the sequence corresponding to this
region is and look across the tree of life, you would see ranging from Bacteria,
Archaea, plants, fruit fly, worm, mouse, human. This sequence would almost be
the same, T-V-G-Y-G, it's very slight variation, and what that tells us is that nature
must have come upon a solution for selective potassium conduction across the
membrane at a very early stage of evolution, and then stuck with that solution, so
now even in our nervous systems we have potassium channels with this particular
structure.

Now I'll show you a movie of this same region, now showing the protein from all
four subunits that would surround the pore to give you a feel for what ions would
experience as they come through. And what is an interesting thing in the crystal
structure here, and that is at one point in the ion conduction pathway there's a
potassium ion with water molecules around it. Now potassium ions typically like to
be in water, and in fact water organizes itself around potassium this way to
stabilize it. And if you look at that and then look in this selectivity filter you can see
what nature is up to in this potassium conductive device, because at each of these
positions—there are four here, where potassium sits—notice that it has four
oxygens from the protein above and four below an ion, and they're organized
around the potassium ion very much the way water organizes around potassium.
And so what nature has done here is come up with a molecule that, in fact, has
sequential binding sites to a potassium ion, and mimic the water molecules around
it. And that's the principle of why the potassium ion comes out of the water and
very happily moves through this device so that you get a selective potassium ion
channel.

Opening and Closing of Ion Channels

Now another feature of these channels to generate electric impulses is that they
can't be open all the time. I mean obviously if all the ion channels in my brain
suddenly opened all at the same time, now my talk would be over and I would be
on the floor and that would be maybe a glorious idea, I don't know, but that would
be the end. But what happens is the opening of these channels is orchestrated in
response to a specific stimulus, and for different kinds of channels that stimulus is
different. Gerry told you about one kind of ion channel, the acetylcholine receptor
channel, that was in the muscle at the synapse so that when the neurotransmitter
acetylcholine was released from the nerve-cell end of the synapse it went and
bound to the acetylcholine receptor and caused that channel to open. Because he
explained that there wasn't enough electricity coming down the impulse to actually
initiate the muscle contraction and it was, in a sense, propagated and amplified by
this synapse. That was a case—the acetylcholine receptor—where the
acetylcholine bound to that kind of channel causing it to undergo a conformational
change to open up. All channels exhibit properties like this. Sometimes it's the
binding of a neurotransmitter like acetylcholine, sometimes it's the binding of an
intracellular ion like calcium, or a G-protein subunit inside the cell, and sometimes it's membrane voltage itself, that charge across the membrane, that does this.

But what's happening when a channel opens and closes? I show this picture here in cartoon version where I have a channel with a little door that's shut, a gate closed, on the inside, and then a door that can open so the ions conduct. So the idea I've been talking about is somehow a signal from within the cell, or from outside the cell, has to cause the door to open and close at the right time. And what an electrophysiologist sees with a very careful measurement of the electrical activity from a very small patch of membrane, and Gerry also showed you this, [is] the picogram-size currents. Here's a case shown here for a potassium channel where when the channel opens the current level jumps up from zero to some value, but what you see here is the channel can open and close, and that would correspond to the gate opening and closing.

So what does this look like in a potassium channel? Well, let me just show you one idea we have from having determined the structure of the case of a closed potassium channel and the case of an open potassium channel. And this is, again, now looking at two subunits of a potassium channel. Again it's the same orientation you've been looking at with the outside on top, inside the cell on the bottom. Here is that selectivity filter without all of its detail, and this is a closed form of a potassium channel. And here is a crystal structure of a different but related potassium channel in its open conformation. It's like opening a door, and in fact if we step back and forth between closed and open we can see that nature is really doing very simple things with molecular machines, almost in the way that we see things in our life. Basically helices, that's what these are called, open up to make the pore wide so ions can cross the selectivity filter, or it closes and occludes it. And so that's a simple kind of conformational change that opens and closes the pore to let the ions flow.

Voltage-Dependent Potassium Channels

But how is this controlled by the cell? That depends on the kind of channel. And now I’m going to tell you a little bit about one kind called voltage-dependent potassium channels, because these are particularly interesting for the nervous system, because it's this kind of channel, a voltage- dependent potassium channel, together with a voltage-dependent sodium channel that actually makes the electrical impulse that travels along the axon of a nerve cell. And a voltage-dependent channel means the following: Here are records of measuring current against time, and you can see there's no channel activity here, but then these two channels in this membrane, one turns on and then another turns on, and they close again. And this is repeated, and what you see are channels open just during a period that corresponds to changing the voltage across the membrane. Now this experiment was done by taking the cell membrane, putting an electrical amplifier across it to control the voltage, step the value of the voltage, and measure the current. And so what you see is that depending on the value of the voltage across
the membrane, or the charge on the membrane, this channel opens up. Now that's interesting actually, because in the beginning I told you that the pumps, remember, separate the ions across the membrane. And then I told you when a channel opens up the ions run down their gradient and they make a voltage across the membrane, they charge the membrane. But now I'm telling you about a special kind of channel that actually, depending on the value of the voltage, determines whether it opens up. So this channel feeds back on itself. It opens and it sets the voltage, but the voltage feeds back and decides whether it should be open, a feedback loop. And it's a feedback loop this way that you need to generate an electrical impulse. So these channels are very important in this way.

The rough picture of how we think they work is shown here. The channel could be closed or it could be open, and notice that in this cartoon I show that when it's closed there are some plus charges on this protein, on this molecule, that actually are near the inside of the membrane, and then when it opens these shift to be close to the outside, and in a sense pull the pore open. Now how would that work? That would work because the voltage across the membrane would affect the positions of these charges. When the membrane's negative on its inside it would draw them inside and close the pore, and when the membrane is positive on its inside it would repel them to the outside and open the pore. So it is, in a sense, a simple mechanism that would occur by the interaction between charges.

Now this thing has a real analogy in electrical systems that engineers and material scientists make that's called the transistor. So the voltage-dependent channels, the voltage-dependent potassium and sodium and another kind, calcium channels, are life's transistors. And transistors are very important in electrical devices because they are conductors that conduct only at certain base voltages and not at others. And these are crucial to all modern electronics. And in fact they're crucial to the electrical system of the nervous system. Now what's really neat here is that I show a graph here from a review some years ago by Fred Sigworth from Yale that shows the relative conductance as a function of base voltage or membrane voltage for the best kind of transistor that material scientists could make. And what this is showing is, in a sense, the turn-on of this device, and this shallow turn-on by the dashed line is an electrical transistor from an electrical device, and the very steep turn-on is a voltage-dependent potassium channel. And so nature has actually made transistors that are much, much better than the best transistors that material scientists can make. Very interesting.

So what do these look like, what do we understand about how they work? This is the picture of a voltage-dependent potassium channel called KvAP from a crystal structure. You're looking down the four-fold, down the pore. This is more complicated than the channel we looked at a few minutes ago, but it is a potassium channel and it has the same pore and the same selectivity filter and the same principles for letting potassium cross, but it's got a bunch of extra stuff attached to that pore. Now if we look more carefully at that stuff, by taking one of the subunits, one of the four subunits out and looking at it from the side in the same kind of
orientation we've been looking at—so this would be outside the cell, and this would be inside the cell, and the pore would be over here. You see a unit, part of this attached extra structure contains a helix-turn-helix element, and it's got some amino acids on it that I show because each of those carries a positive charge. And this is an interesting structure, very unexpected and unusual, and, in fact, we know in the crystallization of this it's not in a completely normal conformation that it would be in the membrane, but what we see is this element that carries charge, and you can do experiments to show that this element moves a long distance across the membrane when the channel opens. And I won't go into those experiments other than to say if we just focused on this element here that's attached to the pore, what we know is that when the channel is closed that element is somewhere closer to the inside of the membrane, and when the pore opens this element is closer to the outside. Now the interesting part here is these shown in gray are the positive-charge-containing amino acids. And remember in the cartoon I showed you a few slides ago I said the general principle by how this ought to work—and it's been understood for a long time—is that when the pore opens it should be moving charge across the membrane, or partway across the membrane. And what we see from these experiments is that they seem to move a long way across the membrane. So this is a conceptual model for how it works, and I call it a conceptual model because for a real atomic-detail model we need structures in both the opened and closed conformations, and even in between, to see exactly how life's transistor moves. But the general idea is that this element, shown now in red—carries the positive charge, and when it would move to the outside it would pull the pore open. And this is the way that the membrane voltage would actually open and close the pore, and therefore this is the way that the channel can set the voltage, but then the voltage feeds back on itself to get an action potential.

**Ion Channels in the Nervous System**

Now the last thing: I just want to end in the last few minutes on an interesting idea, and I don't exactly know what it all means, but the idea is this—I just find it intriguing. Here I am invited to speak at a brain-mind symposium. It's very interesting because I'm a biophysicist who is studying ion channels from Bacteria and Archaea, very simple forms of life. So what do I know about the nervous system? Well in truth not very much. But I was invited here, and actually there's a reason they invited me. It turns out that these simple channels are really very related to the channels that are in our nervous system.

How do we know this? Well, we know this because when we look at the genes of these molecules we can see in the amino acid sequence that they're very, very similar. In other words, these molecules in the nervous system are cousins, or even closer than that, to the molecules that are in some Bacteria and Archaea, simple forms of life. So we know that from the sequence. We know it from the function, when we take these bacterial channels and study them under the conditions similar to the way a neurophysiologist studies the electrical activity in
the nervous system. We see that their properties are very, very similar. So they're really very similar.

And I just want to end by showing you that another reason why we know they're similar, and this reason really tells us how surprisingly similar they are, okay? Surprisingly to me. Here I show a picture of some poisonous animals. You already saw a banded krait, a very deadly snake from Taiwan, in Gerry's talk. And here I show a picture of a . . . this is a green mamba—really deadly—even more deadly I think than the banded krait. The recommendation if you're out in the forest and you get bitten by one of these, and you're very far from antivenom, is find a shady tree and relax. This is a sea anemone, this is a wasp, this is a cobra, this is a scorpion, this is a tarantula. All of these guys make neurotoxins to paralyze the nervous system, and they all work on ion channels.

Now what's interesting is, if you think about something like . . . all of these organisms must have evolved these toxins to paralyze the nervous system of their prey or of something trying to prey on them. And so you'd think that the molecules they have, and the way these molecules work is they're small proteins and they fit like a lock and key, like a key in a lock to the ion channel to plug it. And so it's a very specific shape, designed to basically poison the ion channel. And you'd think that these must have evolved these molecules through natural selection to fit the ion channels of, for example, their prey, that are typically vertebrate organisms that have nervous systems, complex nervous systems. But what happens if you look in the venom of these organisms and ask if these same very specific interactions work against the channels from the bacteria and from the prokaryotic organisms. And what you find is that they do. For example, inside the tail of the scorpion it has venoms that bind with high affinity. They apparently evolved to inhibit eukaryotic ion channels in us, for example, and yet these things inhibit very specifically the ion channels, even in bacteria.

Another example—a tarantula from Chile. This guy specializes in a toxin that binds to that voltage sensor, and apparently that evolved to recognize the voltage sensor in organisms with complicated nervous systems, and yet these same proteins fit very specifically to the voltage sensor of the bacterial channels. That's very interesting, because it tells us that not only are these channels in the prokaryotes related, but it tells us that they're really related. They've maintained their structure over the eons of evolution, they've maintained the detailed shape.

And I end with this picture that I like to think about, because it strikes me, because here, for example, this tarantula is from the dry scrublands of Chile. It makes a poison to inhibit the nervous system of other insects, for example, and yet we find that the toxins inside his venom actually recognized the channels in an Archaea from thermal vents in the Japan Sea. Isn't that amazing? That's really crossing unusual geographic and traditional evolutionary boundaries. But what it's really telling us is that there's a high degree of molecular conservation. And what that seems to tell me is that these molecules that are in our nervous system that are
doing the fancy electrical signaling that is required for us to move and think actually must've evolved for other purposes. And then it must be that when the evolution of needing a complex nervous system for electrical communication—it seems to have operated on what was available, because certainly this bacterium doesn't have a nervous system. And it wasn't repolarizing its action potential with a potassium channel. So I just want to end on the idea that it's fascinating that we can look in the simple organisms and find out so much detail about the hardware of the nervous system in complex organisms such as ourselves.

And finally I just want to end with an important acknowledgment slide for the people who have worked so hard with me over the years on the projects that I talked about, Joao Morais Cabral and Yufeng Zhou, who are both in different departments at Yale now. Youxing Jiang, who is at University of Texas Southwestern, also Amelia Kaufmann, Alice Lee (MacKinnon) and Jiayun Chen, and Vanessa Ruta, a student in my lab, who did many experiments on the voltage-dependent channel, collaborator Brian Chait at Rockefeller University, and his post-doctoral scientist Martine Cadene. Thank you very much.

Questions and Answers

**Man:** On the concept of molecular machines, real theoreticians use the concept of dynamic self-regulation in organisms. What's the difference between yours and theirs? The organism generally is not considered a machine.

**MacKinnon:** So the question was, I used the term molecular machine and then it was brought up that scientists in other fields talk about dynamic self-regulation and what's the relationship? And I'll give a try at it, very simply that in a sense I think dynamic self-regulation would refer to—I guess would be a very general statement—about the way a system can, in fact, regulate itself. Whether the system contains many components like you or me, or maybe even a society, or a whole planet, or down to something as simple as a molecule. And I guess the simple relation... thing I could connect in your question is that this particular machine, the voltage-dependent channel, is really, in a sense, compared to you and me, is a very simple thing. And yet even it feeds back and self-regulates itself. It self-regulates its shape, in a sense—whether its gate is opened or closed.

**Man:** [Inaudible.]

**MacKinnon:** We could talk afterwards. That would be nice. Thank you.

[Question inaudible.]

Yes. The question is, Is there a pH dependence? And yes, like most proteins, its opening and closing especially is sensitive—its conformational shape is sensitive to pH, some of them more than others. The voltage-dependent channels are quite sensitive to the pH, and in fact there are some channels where actually the pH is
probably the main regulator that causes them to open and close. So there are cases like that, yes.

[Question inaudible.] 

So the question was, How would I compare the structure of this potassium channel to really much simpler molecules like valinomycin? Valinomycin has not many atoms at all and yet it selectively binds potassium. And in fact . . . And the question is, Could there be an evolutionary connection? It turns out if you look at valinomycin’s structure with potassium bound to it you realize the basic chemical principle used by valinomycin to selectively bind to potassium is the same as in the potassium selectivity filter in the way it coordinates the potassium ion with oxygens. But by comparing the structures you can also tell that, in a sense, it's using the chemical and physical principles but it's actually not evolutionarily related, so in a sense it's an independent solution to the problem, but the same basic physical properties.

Thanks.
Introduction by Gerald Fischbach

Gerald Fischbach: Thanks, Rod. One of the groups that Rod thanked was the National Institutes of Health, and I want to put in a plug for this, for everyone here, understanding the huge role that public funding for our research has played, and how now at a time of leveling-off of the funding I hope everyone in this room will realize the need to support the NIH. To support fundamental science in this country, and in the singularities of investigation like you just heard, basic research that will improve the health of this country. We should know about spiders and toxins if we're going to understand epilepsy and thought disorders and a number of others. So I want to add my thanks to the NIH.

Tom Jessell, I believe, is the leader in developmental neural science in this country and around the world today. He's been interested in motor neurons through much of his career, almost all of his career, and he has focused on the embryo much like Ramón y Cajal focused on the embryo to simplify things and to understand how neural circuits get assembled. It was thrilling a week ago to hear Tom take part in this symposium, which really originated with studies done in the '50s and '60s . . . fundamental questions of how one nerve cell recognizes another and forms a synapse on it. To see the culmination of this work in the molecular studies Tom will describe in synapse formation within the spinal cord. And you will see that his work has immediate implications for devastating diseases of motor neurons, including amyotrophic lateral sclerosis and spinal muscular atrophy. Tom also has a number of very well-deserved honors, being a member of the Royal Society of London, and as associate of the National Academy of Sciences. Very few people with foreign citizenship are elected to our national academy, and Tom is one of them. Tom?

Neural Circuits and Behavior

Thomas M. Jessell: What Gerry didn’t mention is that I was a student of Dr. Fischbach 25 years ago, and if he was in a less modest mood he would tell you—and who's to deny it?—that he taught me everything I know. And it's a pleasure actually to be reunited with Gerry in his new position as dean as Columbia moves into an exciting new phase in the neurosciences.

So human beings, as with all other forms of animal life, display a remarkable capacity to interact with and respond to events in the world that surrounds them. These actions and reactions to a large extent underlie and define our behaviors.
And the repertoire of behaviors that is exhibited by an individual organism, a single animal, is extraordinarily diverse. And some of these behaviors are relatively simple. The ability to move and to act in response to sensory stimuli, and others are more complicated and involve cognitive processes and influence mood, affect, and thought. But each of these behaviors, whether simple or complex, depend[s] on the functions of nervous systems, and in particular depend[s] on sets of neurons that form connections with each other, so-called neural circuits.

And so one of the challenges in neural science is to understand the relationship between the organization of neural circuits and the emergence of behavior. And over the last fifty years or so, in large part through clinical studies, we've come to understand that circuits in different regions of the brain are assigned to different specific behavioral functions. And one striking example of that was shown by Gerry in the use of deep-brain stimulation to affect motor and movement disorders. But there are many, many examples of this sort.

But despite this progress, I think it's reasonable to say that we have still a very poor and unsatisfying understanding of the relationship between the detailed workings of the neural circuit and the emergence of the behavior from that circuit. So in many ways one of the challenges that faces neural science today is shown in this first slide here. And so if [what] one is trying to do is establish the link between the properties of individual neurons, the way they assemble into circuits, those circuits are embedded in the brain to produce certain specific and idiosyncratic behaviors.

And in brief, if we understood how circuits in a sense accounted for the essence of Eric Kandel then the problem of neural science would be solved and we could all go home. But unfortunately we don't understand this link. And so one of the challenges is to try and make that link between circuitry and behavior. And there are many ways of approaching this problem, but being a reductionist what I'm going to try and do today in this morning's discussion is really make an argument that perhaps we can gain insights into the link between circuitry and behavior by studying how these circuits are assembled during embryonic and postnatal development.

If we understand the organization of circuits during development, perhaps we can understand and glean new principles about the way those circuits function in the mature state. And a better understanding of the structure of these circuits may give us insights into the way that behavior is reflected in changes in circuitry during experience, learning, and in disease states.
Assembly of Neural Circuits

So what do we know at the moment about the organization and assembly of neural circuits? Work over the last thirty to forty years has told us, I think, that the assembly of a circuit occurs in progressive stages with, perhaps, two major components. The first is simply by observation we've learned that circuits are assembled from embryo to embryo from organism to organism in a highly stereotyped reproducible fashion. Neurons don't grow out at random to innervate any number of potential targets, they begin to develop in a highly organized way. And I think this is interesting because it essentially makes an argument for a hardwired state of neural circuit formation. And the implication of that is that perhaps there are genes and genetic programs that underlie that hardwiring that leads to the first aspect of circuit formation.

But the final pattern of circuits that we see in the central nervous system, in the brain and spinal cord, is simply not controlled by these genetic hardwired programs, but then the experience, the environment that an organism [is] exposed to then, in an important way, refines and modifies those circuits. So one can think of development as first initially being a genetic program, which then gets later modified by experience and activity.

But what I'd like to do today is try and point out that in trying to understand the progression of circuit assembly perhaps the important key, at least at a reductionist level, is the idea that the driving force for circuit assembly is the assignment or the allocation of neurons with particular identities. Now neurons cherish their identities because what the identity of a neuron is doing in early stages of development is, for example, defining where those neurons are positioned within the brain, and by implication who their neighbors are. It's defining how neurons are capable of communicating with other classes of neurons through the process of extension of axons from the site of generation to distant target regions, and then neuronal identity also determines the formation of stable and meaningful interrelationships with other classes of nerve cells through the process of synapse formation that Gerry has indicated.

So the brain contains hundreds, perhaps thousands, of different classes of nerve cells, and probably an equivalent number of neural circuits. So in trying to approach the problem of the assembly of a circuit, one could choose, and people are choosing, many, many different regions to study this general issue. What I'd like to do this morning is to illustrate the way in which genes, neurons, and circuits might influence behavior by focusing on one particular set of behaviors, motor behaviors, the control of movement. Because if one thinks about it, motor systems, the control of movement, is in fact the manifestation, the expression of all aspects of animal behavior. Without movement, despite what you think, there's no way of expressing your behavior to others. And one particularly graphic way of bringing this point home deals within a clinical context: patients, for example, who've had cerebral infarcts or strokes who preserve mental and cognitive functions but have
no ability to communicate that information, their thoughts, with the outside world. And Jean Bauby some time ago wrote a very graphic description of what it feels like in this book *The Diving Bell and the Butterfly*, where after a stroke he was paralyzed and had the inability to communicate in any way other than moving his left eyelid. So this illustrates the importance of motor systems. And the central nervous system commands its motor systems through circuits that reside within the spinal cord.

And the spinal cord is an interesting region of the nervous system to study general problems of circuit formation for a couple of reasons: The first is that one of the big problems, how sensory information is transformed into motor output, is occurring in a relatively immediate way in the spinal cord compared to many other circuits that are found, for example, in the brain. But the spinal cord also gives one, in a sense, a nice reflection of this interplay between genes and experience in the control of circuitry. No one who has observed a newborn calf walking within minutes after birth I think could argue against the idea that there must be innate programs of circuitry that control movement in the absence of learned experience. On the other hand, the ability to play a musical instrument clearly indicates that the motor system is acquiring learned behaviors. So perhaps by studying circuits in the motor system one will begin to be able to get insights into this balance.

And within the spinal cord one class of nerve cells, the motor neurons, which Gerry has introduced to you, play an essential role in communicating the central nervous system's interaction with the periphery, with target muscles. And one of the motor functions performed by the spinal cord that is particularly evident is that of locomotion, the ability of animals to translocate from one position to another. And a somewhat fanciful view of motor control but which illustrates some of the points I want to refer to is shown in this slide from the Italian futurist Giacomo Balla, and so this is the dynamism of a dog on a leash. But what I want you to focus on is the coordinated limb movement that occurs in a conserved manner across mammalian evolution. And so in order to take a single step or to walk or to run, several things have to happen in terms of motor control, muscle coordination. So, first of all, individual muscles in the limb have to be activated in a precise choreographed manner in order to take even a single step, and then the activity of programs of the limb, motor programs, on the left and right sides of the body have to be coordinated in order to achieve this locomotive behavior.

**Diagram of a Core Motor Circuit**

So what I'd like to do this morning is really take you into the workings of the central nervous system from a developmental perspective and see to what extent genes can explain neuronal identity and how that drives the circuits that control motor behavior. And so in order to do this, we're going to look now at a diagrammatic view of a core motor circuit that resides within the developing and mature spinal cord. And that circuit is shown here. So now this is a cross-section through the spinal cord, through your spinal cord, through any vertebrate's spinal cord, and
what we can see here are the positions of motor neurons, which as Gerry showed are located in their cell bodies within the spinal cord and then they send axons out into the periphery over long distances to reach individual muscle targets. So some motor neurons are dedicated to the control of extensor muscles, other motor neurons are dedicated to the control of flexor muscles. But motor function cannot exist through motor neurons alone, there have to be important additional neuronal elements that control motor output, control motor behaviors. And I'm going to be discussing two classes of neurons that impinge and influence the activity of motor neurons.

So the first are sensory neurons, and the job of the sensory system is to convey the state of muscle contraction and relay that information from the periphery back into the central nervous system. So sensory neurons send axons, which monitor the state of muscle contraction, project via a second set of axons into the spinal cord, and in some cases this sensory motor transformation is remarkably direct through the formation of direct monosynaptic connections between sensory and motor neurons. And it's these connections that drive the so-called monosynaptic or knee-jerk reflex, which all of you can experience as the activity of this particular synapse in the central nervous system. So sensory feedback is important in fine motor control, but then in addition the phasic firing properties of motor neurons that are involved in the locomotive behaviors we've seen depend in addition on sets of interneurons that are actually buried close to motor neurons within the spinal cord. So there are sets of interneurons allocated to the distinction in firing of motor neurons that project to extensor and flexor, and also ensure the coordination of motor output patterns on the left and right side of the body.

So how, in a sense, does this circuit arise during development? So in order to approach this issue we have to go back from the details of the mature circuit and move back into embryonic development and begin to address how this circuit emerges. This picture is actually a real image of a developing mouse embryo about midway through gestation in which motor neurons located in the ventral spinal cord are shown in green and sensory neurons in red and blue. And the main point here is to illustrate that even at this early stage, many of the cardinal features of motor innervation of muscle have been established. Motor axons have emerged from the central nervous system, begun to project into different peripheral target sites, and the question of motor neuron identity can be inferred in part from looking at the different trajectories of the axons, these green processes here, of different sets of motor neurons.

So as we approach the problem of motor neuron identity how can one begin to organize in some practical form? And so here is a diagrammatic view of the spinal cord, which illustrates some of the challenges that motor neurons have to undertake in terms of acquiring identities that determine their target projections. So this is looking along the rostrocaudal axis of the spinal cord, the head-to-tail axis here, so some motor neurons are formed at limb levels, those neurons will
innervate muscles in the limb; others are formed at trunk, at thoracic levels, and will innervate different motor neuron targets.

So one can think about the problem of allocation of motor neuron identity in several subsequent steps, progressive steps. So first of all the developing embryo has to decide to make a motor neuron as opposed to the many hundreds of other classes of neurons that are found in the central nervous system. Because of those thousands of neurons it's only motor neurons that have the capacity to extend axons out of the central nervous system and communicate with the periphery. And then if we're trying to innervate limb muscles to control locomotion, motor neurons have to acquire a limb identity that ensures that they innervate the right peripheral targets. And then each individual muscle is controlled by one set of motor neurons, so-called motor neuron pools. And so we need to understand the identity of this fine degree of motor neuron diversity because that is the core unit, in a sense, of this circuit. Motor neuron pools innervate muscle targets. They receive selective sensory feedback information, and they also drive the specificity of interneuron connections.

**Signals Establishing Cell Identity**

So this is already looking rather complicated, but what I'd like to try and show you is that, in fact, one key developmental principle addresses or is relevant to each of these steps of neuronal diversification, of neuronal identity. And this is a development principle that applies not just within the nervous system but of the shaping of tissues and organs throughout developing embryos. And that molecular principle is shown in this slide here. And so this could be a neuron or a neural precursor cell, a neural stem cell, and this cell acquires its identity through a relatively simple molecular program. Its identity is influenced by signals to which it's exposed at early stages of development, often protein factors that change the fate of that cell, divert it from one potential fate to an alternative fate. And the way that these extracellular extrinsic signals signal to this neuron or this neural precursor cell is typically throughout transmembrane receptors, similar but importantly different from the sorts of transmembrane proteins that Rod has described to you. That signaling pathway is then transferred into the nucleus of that cell where it induces the expression of a set of genes called transcription factors, and these are proteins that have the capacity, once expressed, to bind to DNA elements with high specificity and to induce the expression of certain target genes. And these target genes in this context are the proteins that drive neuronal circuitry. So one can think of this developmental principle in two important steps, how extrinsic inductive signals control transcription factor expression, [and] how these transcription factors direct the downstream expression of proteins that control neuronal circuits. And so I'm going to try and illustrate how these two principles shape motor neuron projections, shape sensory innervation into motor neurons, and control interneuron diversity that influences motor output.
So the inductive signals, if we begin there, within the spinal cord, operate along two major axes within the developing embryo. So if this is the early spinal cord, you can see that the spinal cord is positioned within the embryo along two axes. One is the dorsoventral axis, the top-to-bottom axis here, and a second is an anterior-posterior, the head-to-tail axis. So the position of a motor neuron, in a sense, can be described by its position within these two coordinate axes. And different signaling factors, different inductive signals, contribute in a collective way to assigning motor neuron identity. And I'll just illustrate what we currently understand about the way these inductive signals control motor neuron identity.

So the first set of pathways I'll talk about along the dorsoventral axis, so here is a cross-section through the early spinal cord, the neural epithelium, and you can see individual neuroepithelial cells before they've made the decision to give rise to motor neurons. And at this stage, all neuroepithelial cells are essentially identical. And they acquire their different identities in part through their exposure to different environmental signals. And the key signal that controls motor neuron and different classes of interneuron identities is a secreted protein known as sonic hedgehog, which derives from these two ventral midline structures. It has this bizarre name because it has a counterpart in fly patterning in embryonic development, and fly geneticists have a propensity to naming the genes that affect developmental processes in rather more exotic ways than their mammalian geneticist counterparts. But nevertheless sonic hedgehog is secreted from these two cell types, and establishes a protein gradient within the neural epithelium so that the position of a cell determines its identity by virtue of exposure to this concentration gradient of extrinsic sonic hedgehog, ensuring that motor neurons generate in one position and different classes of ventral interneurons that contribute to the motor circuit are generated in adjacent spatial domains.

So in turn this sort of observation raises the issue of how a single cell can respond to a gradient of this extrinsic signal and acquire an all-or-none identity, because these cells make up their mind with unerring precision. So how is that step achieved? It turns out that, as I mentioned, the way in which this extrinsic signal establishes cell identity is by setting up spatial domains of transcription factor expression, and along this axis it turns out to be a set of homeodomain transcription factors. But the main point is that different dorsoventral domains have different molecular identities revealed by differences in transcription factor expression, and that an individual cell is forced to choose one or other of two alternate identities in a winner-take-all strategy because these transcription factors talk to each other within the cell, so one transcription factor cannot coexist in the presence of an alternative transcription factor through repressive interactions. These cells force an identity on a progenitor cell to give rise to a motor neuron as opposed to an interneuron.

So this gets you to a generic motor neuron state, but how do motor neurons, for example, destined to innervate the limb, acquire their particular limb level identity? And a very similar developmental principle operates along the dorsoventral axis to
the one that I’ve just described along the rostrocaudal axis. So here there is another graded signaling factor. It turns out to be a member of the FGF class, fibroblast growth factor class of secreted proteins, which establishes a gradient along the anterior-posterior axis. That gradient is then read out, as Jeremy Dasen’s work in the lab—together with Serena Liu—has shown, in the patterned expression of Hox proteins, another set of homeodomain transcription factors. So limb-level motor neurons express one Hox gene, shown in green, thoracic level express a different Hox protein. And again this intracellular fight between transcription factors ensures a unique identity. So in this way inductive signals in the environment gradually induce programs of transcription factor expression that lock cells into particular neuronal, in this case motor neuron, states.

The other remarkable thing about this patterning mechanism is it is not the particular consign of the spinal cord in a vertebrate embryo. The Hox genes were first discovered for their ability to pattern the entire embryo in insects, such as a Drosophila embryo here. And in fact the same is true for Hox genes within the mammalian embryo. So it turns out that Hox genes influence not only where motor neurons are generated but the position of limb formation, thus ensuring that the right set of neurons and the right target are organized in register in the developing embryo. So through this progressive set of interactions we can begin to see how extrinsic signaling factors control transcription factor expression and gradually allocate neurons with particular identities.

**Transcription Factors Controlling Innervation**

And the same is true when we move further along to the pool identity. It turns out that different Hox genes correlate with particular motor neuron pools and that these Hox genes then drive the expression of yet further classes of transcription factors. So here we’re now going to move from the first of these developmental principles about [how] extrinsic signals control transcription factor expression, and now begin to look at how these transcription factors that are allocated to particular motor neuron pools in fact control later aspects of motor neuron development, the ability of a set of motor neurons to innervate on target muscle and not another.

And one way of thinking about that is just to focus on what are the tasks of a single set of motor neurons as they emerge from the spinal cord and project to their muscle targets. And we can think about this in three sequential steps: First of all motor neurons have to get out of the spinal cord, but that is not good enough, they have to know to innervate one target muscle, an extensor muscle for example, and ignore the ability to innervate a different functionally antagonistic muscle. So this is a process of axon pathfinding and target selection. Then once the axon reaches the vicinity of the muscle it has to know to branch and form effective synapses with that target muscle. And then in addition motor neurons, although initially scattered within the spinal cord, cluster together into so-called motor neuron pools, and that the reason they cluster is that this allows motor neurons to communicate with each other through ion channels of a family related but different to those that Rod
MacKinnon has talked to, but these ion channel communication steps ensure that all the motor neurons that project to a muscle fire in a phasic coordinated manner, which is important for locomotion.

So it turns out that these different transcriptional modules that have been defined in motor pools have different jobs. The process of axon pathfinding, reaching the muscle target, is the responsibility of one set of transcription factors, the Nkx6 proteins, as Natalia de Marco's work has shown, but that these later steps, the branching and innervation and the clustering of motor neurons into pools actually are the job of the ETS transcription factors. And just to illustrate how transcriptional identity can drive later aspects of circuitry, I'm going to focus on this one class of transcription factors of DNA binding proteins.

So first of all I'll just show you some evidence that the innervation and the clustering of neurons depend on these transcription factors, and we can get this evidence simply by eliminating through genetic means these transcription factors from mice and examining the development of motor axon projections. And so this is looking at the mice, which are either normal or mutant in a particular ETS gene, focusing on the branching. This is one muscle target. We've used a genetic trick to light up motor axons in green, and you can see a perfect innervation of both the proximal and distal regions of this muscle in the presence of this ETS transcription factor. In the absence of this gene what you can see is that vast regions of the muscle are deinnervated here. So it's this type of evidence that indicates that late aspects of innervation of muscle target depend on this set of transcription factors. It turns out that the same gene that is controlling muscle branching in the periphery is controlling the clustering of motor neurons in the spinal cord, as work by Silvia Arber and Jonathan Lin and Chris Henderson showed. So here is one of these so-called motor neuron pools, a tight cluster of motor neurons within the sea of motor neurons that innervates the limb. If you eliminate this one ETS gene then instead of being tightly clustered these motor neurons are now scattered at random throughout this set of motor neurons. So in this way then transcription factors not only define terminal branching but also the organization of neurons within a motor pool.

But these are transcription factors, these aren't the proteins that really [are] the workhorse that are controlling these developmental events. So presumably the transcription factors are, in turn, controlling cell-surface proteins that recognize the environment and mediate these different motor neuron behaviors. And in some cases we know the targets of these transcription factors that actually control these behavioral properties. So I'll just give you one example, which is how ETS transcription factors ensure motor neuron clustering. And it turns out that these transcription factors control a set of cell-surface recognition proteins called the cadherins. And so the cadherins, from work by many groups, has revealed that what these proteins, which are sitting in the cell surface membrane, like to do is bind cells together, and they bind cells through so-called homophilic, like-recognizing-like reaction. And what that means is if you have two groups of cells,
which express different cadherins on their cell surface, even though initially mixed, these two cell populations, through cadherin recognition, will sort into distinct clusters. And that is, in fact, what is going on in terms of the role of cadherins as we think in motor neuron sorting, so the selective expression of cadherins under the control of ETS proteins by one and not another set of motor neurons ensures that they cluster together to form discrete pools.

So that in this way genes, transcription factors that are regulated by environmental signals and control cell surface proteins, ensure that particular sets of motor neurons are organized within the central nervous system and innervate different target muscles in the periphery.

**From Individual Neurons to Circuits**

So how can this transcriptional logic move from the properties of an individual neuron to begin to reconstruct a circuit? So I want to begin to talk now, very briefly, about the role of sensory feedback information from primary sensory neurons. The job of these sensory neurons is to relay information from the periphery to innervate particular sets of motor neurons in the central nervous system. So what sensory neurons are doing is giving the motor system, if you like, a sense of place, it's telling the central nervous system about limb position. So whether you're Michael Jordan or Margot Fontaine the ability to control one's limbs in a precise pattern depends not only on the motor system but depends on sensory feedback information that is monitoring limb position and the state of muscle contraction.

So one of the remarkable things is that the transcription factors that I've described to you, the ETS proteins that assign motor neuron identity also are expressed by the sensory neurons that provide this feedback information. So that it turns out that not only are these genes, these red and yellow nuclei or red and green nuclei, here defining individual sets of motor neurons, but if we look in their position of the sensory cell bodies, they're defining sets of sensory neurons, so that introduces the idea that, in fact, elements of this circuit that are going to functionally communicate sensory and motor neurons are defined genetically by the matching expression of one set of transcription factors. And we know that these transcription factors are important not only for motor neurons for also for sensory neurons, because again if we use mouse genetics to eliminate these genes we find dramatic defects in the ability of sensory neurons to communicate with motor neurons. And I'll just show you one example from a mouse mutation in another of these ETS genes. So in diagrammatic form what the sensory neuron has to do is grow into the spinal cord, grow down into the ventral region, and find and form synaptic connections with motor neurons. Here are these sensory projections, and here is the synaptic potential recorded physiologically. If we get rid of this ETS gene, then the sensory neurons are still there. They can make it halfway into the spinal cord, but they completely fail to project into the vicinity of motor neurons, and not surprisingly in the absence of this contact there is no synaptic communication between these two cell types.
So this begins to suggest that transcription factors are good not only at controlling neuronal identity but beginning to build up aspects of circuit formation. The problem doesn't stop for the sensory neuron in just innervating the motor neuron, you remember it has to innervate the right type of motor neuron in order to produce functional output. And so what this system offers an opportunity to approach is the problem of synaptic specificity, why some sensory neurons innervate some motor neurons, ignoring others. And with new anatomical methods developed by Julia Kaltschmidt, we can actually see a motor neuron and can see the synaptic terminals in this region on a motor neuron. So we need to understand how this particular set of sensory synapses chooses this motor neuron, ignoring another motor neuron. And again these transcriptional cassettes that I've been describing may provide some insight into this problem. Because it turns out that not only are the ETS genes expressed by sensory and motor neurons, but the cadherin proteins, which we know are targets of the ETS genes in the motor system, are also expressed by the sensory neurons, and there is a correlation between gene expression by sensory and motor.

So this leads to the intriguing, although still untested, idea that perhaps recognition between a sensory axon terminal and the motor neuron dendrite that accounts for the functional specificity in the system is mediated by these cadherin homophilic recognition interactions. And so genetics should allow us to probe this circuitry in more detail.

**Interneurons and Control of Motor Output**

So this gets us from a sensory to a motor system, but now for the last few minutes what I'll move to is the important role of local circuit interneurons in driving the phasic firing patterns of motor neurons. And we can think of these spinal interneurons in two sets, those that are involved on one side of the spinal cord in controlling flex and extensor movements, and those that communicate information across the spinal cord, ensuring the left-right phasing of motor output. And one graphic way of illustrating how these interneurons control motor output comes from actually watching the activity of motor neurons in the spinal cord as an animal steps during a walking movement. And so what we're looking at now is a top-down view of the spinal cord where each of these colors is actually a cluster of motor neurons, the so-called motor neuron pool. And the colors indicate the state of activity, of firing of these motor neurons—blue, cold colors, no activity, warmer, red colors a high level of activity. This is a static image. What happens if an animal begins to walk, what happens to the firing patterns of these motor neurons? And what I think you can see here is that the firing patterns change dramatically but in an organized way. You can see that motor neurons on the left and right side are never firing in phase, and you can see that motor neurons at different levels of the spinal cord are also firing out of phase, which corresponds to the innervation of flexor and extensor motor neurons.
So can we use this information and begin to pick apart through genetics the ability of interneurons selectively to control, for example, left-right coordination or extensor-flexor coordination? And in studies performed together with Martyn Goulding's lab at the Salk Institute, we've taken advantage of the fact that in just the same way that motor neurons have a transcriptional identity, interneurons have a transcriptional identity. So mutating the genes that control interneuron identity should permit us in a way that is not easily possible through other routes to selectively inactivate or eliminate one of these sets of interneurons and examine the consequences for coordinated motor behavior. So there are two classes of interneurons, as I mentioned, the so-called V0 interneurons, which mediate communication across the spinal cord, and a different set involved in flexor-extensor. One of the remarkable things about this motor activity is it doesn't require sensory feedback, it doesn't require descending information, and you can isolate the spinal cord and see these patterns of phasic motor activity. So here we're recording from the ventral route the axons of motor neurons, a burst of action potentials, from a set of motor neurons that innervate a flexor muscle. And the important thing, if we concentrate on this burst is that, for example, on the right-hand side if we look on the left-hand side the bursting is perfectly out of phase. Similarly flexor and extensor bursting shows alternating phases. And this is completely in vitro, implying that these local circuits operate in the absence of sensory feedback information. So this is the normal situation.

What happens if, for example, we eliminate this set of interneurons through genetics, what happens to locomotive behavior? Then in the absence on these crossed interneurons what you find now is that the left-right normal phasing is completely degraded and now motor neurons on the left and right side of the body fire in bursting, so you've eliminated the left-right coordination while preserving the flexor-extensor coordination, which is presumably the function of this set of interneurons. So in this way we can begin to dissect out through genetics local circuits as well as sensory feedback as well as the motor innervation of muscles.

Unresolved Issues

So what I've tried to do is to indicate in this very simple system a core circuit that controls many aspects of motor behavior, involving motor neurons, sensory neurons, and local circuit interneurons. And at least from our perspective the key to understanding the development of this circuit goes back to the simple view that inductive signals control neuronal identity, neuronal identity is established by transcription factors, which bind to DNA targets and induce target proteins. And these target proteins, as I've mentioned, control neuronal position, axonal trajectory, the formation of synapses, and that collectively this is what constitutes a neural circuit. And these neural circuits, as we've seen, control simple motor behaviors.

But there are several unresolved issues, I think, that relate to this sort of even simple circuitry. First of all I've described a core circuitry that controls vertebrate
locomotion. But in fact if you look across vertebrate organisms, and there seems to be a consistency of showing slides of snakes in this symposium so far, what you can see is that different vertebrates have evolved different ways of using this core motor circuitry to suit their particular behavioral requirements. And it seems likely that these different adaptations of this core motor circuit are in fact innate, are genetically programmed, so one of the things that genetics and circuitry has to try and do is understand how you modify this core genetic circuit to produce the specialized behaviors that different vertebrate organisms display.

And the second challenge is to come back to this issue that so far what we’ve been talking about is solely this early hardwiring component, and upon that we have to understand how experience and how activity and how interaction with the outside world modifies this core circuitry.

So to end as we began with a slide from Balla, the hand of the violinist, here you can see that this fine degree of motor coordination is not acquired de novo, has to be learned. Where are the circuits that control that learning process? We know that some of those circuits exist within the central nervous system in the brain. Are there changes in circuitry in the spinal cord that also contribute to these learned motor behaviors? So there are many challenges, but nevertheless the hope is that by understanding the details of circuitry in this one system we can extract principles that are going to be relevant to the formation of circuits in many other regions of the brain that control much more complex diverse behaviors. And I think the next ten to twenty years is going to see an increasing ability to meet those challenges.

And I will stop there, but again just emphasize the people in my lab who’ve done much of the work that I’ve summarized here. The work on transcriptional identity and motor neurons is that of Jeremy Dasen and Serena Liu. Much of the work on motor axon outgrowth and synapse formation is from Natalia de Marco, [and] Ivo Lieberam. The work on sensory motor synapses is that of Julia Kaltschmidt. And the work on these ETS transcription factors and cadherins are that of Silvia Arber, Jonathan Lin, [and] Stephen Price. And the work on local interneuron control is that of Alessandra Pierani. And we’ve enjoyed many collaborations with other groups during the course of this program.

Thank you very much.

**Question and Answer**

I’ve been informed that there’s time for a couple of questions, if people have burning issues.

[Question inaudible.]
Yes, I'm going to rephrase the question. Are there other proteins that might be involved in these developmental processes related to motor neurons? And the two that were mentioned were particular proteins, neuregulin and agrin. What I've given you is really the tip of an iceberg here of a very complicated molecular program, so this is not to say that these are the only genes that are going to be controlling aspects of sensory motor development. And one of the tasks is to work out how this complicated array of genes fits together, and I think this is one of the things for the future.
Brain and Mind
May 13, 2004

Richard Axel, MD
Scents and Sensibility: Towards a Molecular Logic of Perception

Introduction by Gerald Fischbach

Gerald Fischbach: Richard Axel is a University professor at Columbia and was, seems to me, born and educated at Columbia, has been an extraordinary influence not just in neuroscience but I think in American biomedical science. He had a career in molecular biology—and still does—that has enormous influence over the way we think about genes and gene expression, and it is the infusion of new talent—from molecular science, from physics, from chemistry—into neuroscience that has invigorated this field. If you really think about some of the advances and some of the people speaking today, you will see that people are joining this field from very different walks, I think because of the challenges that the brain presents, both from the molecular point of view and from the more integrative behavioral point of view.

Richard had a career in medicine, and to illustrate how creative and innovative he is, when he got his degree in medicine, they promised they would give him the degree only on the condition that he never touched a live patient. So he switched fields, and he’ll tell you the rest of that story. But he switched fields to our great, great benefit. Richard has won many, many awards. A few years ago his laboratory made a startling discovery, one of the most revolutionary discoveries of our time: that there is a gene pool, a very large gene pool, perhaps the largest pool of genes in the genome, that encode receptors in the nose that detect odors and olfaction. This was a triumph that is really hard to describe in terms of the impact on the field of sensory biology. But as he will describe, it promises to go far beyond that to the point of really trying to understand how we deal with sensory information, how we perceive it, and what the meaning of this is for the functioning human organism.

Perceptions are Internal Constructs

Richard Axel: Thank you, Gerry. As Gerry pointed out, indeed, I was afforded an MD with the promise that I never, ever practice medicine on live patients. And I kept my promise and returned to Columbia in the pathology department here where I did a year of pathology and afterward the chair of pathology offered board certification if I promised never, ever to practice medicine on dead patients. And indeed this led to my current endeavors, and for this and for many other things I owe this university an enormous debt of gratitude. I love this place.
This is not a nose. It is a portrayal by the Belgian surrealist René Magritte of his own brain’s representation of the external world. It is a vignette of image and reality locked in mutual consolation. The sense of slippage between image and object is a source of creativity persistent in art brought to its culmination by the surrealists. The problem as to how the brain represents the external world is not only at the center of modern art, but it is at the very core of philosophy, psychology, and neuroscience. All organisms have evolved a mechanism to recognize sensory information in the environment and transmit this information to the brain where it then must be processed to create an internal representation, a map of the external world. And the existence of a map in the brain immediately implies that different species—and at the extreme even different individuals within a species—will represent the world in different ways. Indeed studies on the evolution of sensory perception reveal the important fact that each species lives in its own unique sensory world of which other species may be partially or totally unaware.

Bats, for example, extract remarkably detailed information about the position, velocity, and size of objects in their surroundings from biosonar, or echo location. Some fish use conceptually very similar modalities that involve electromagnetic waves to navigate and detect prey. Snakes, here you go, boas, pythons, pit vipers shown here, maintain highly sensitive infrared imaging systems that target prey in the absence of visual information. These snakes have evolved a highly specialized sense apparatus, the pit organ, which detects heat from emitted infrared irradiations. These sixth, seventh, and eighth senses, which humans do not possess, illustrate quite clearly that each species perceives but a meager image of the richness of the outside world. The brain functions, then, not by recording an exact image but by creating its own selective picture. Our perceptions are not direct recordings of the world around us; rather, they are constructed internally according to innate rules. Colors, tones, tastes, smells are active constructs created by our brain out of sensory experience. They do not exist, I argue, outside of the brain. That which an organism can perceive is determined by the unique allotment of neurons with which it is genetically endowed, and we are therefore trapped in the representation of the world made possible by our genes.

But how can genes provide insight into the astonishing problem of how the brain represents the outside world? The brain, after all, consists solely of a collection of excitable neurons. And how is it that the rich array of mechanical, optical, and chemical properties that define touch, hearing, vision, smell, taste can be represented by bits of electrical activity that can essentially only vary in two parameters, time and space?
Odor Recognition and Discrimination

Now we are interested in how it is that olfactory information is represented in the brain. In humans smell is often viewed as an aesthetic sense, as a sense capable of eliciting enduring thoughts and memories. But smell is a primal sense. For most organisms it is that sense that affords them the ability to detect food, predators, and mates, and evolutionarily it is the most primitive sense. What we wish to determine is how brain space represents chemical structure, how the brain knows what the nose is smelling.

Let's consider this in the context of the anatomy of the nose. The olfactory sensory neurons reside within a sheet in the posterior recess of the nose. The neurons themselves are very simple structures: they're bipolar. Shown schematically, here the individual neurons send a process, a dendrite, out to the surface of the nose in contact with the external environment, and it is on these processes that reside the receptor molecules capable of interacting with the universe of odors. The energy of binding of odors to these receptors is transduced into electrical activity, which then travels down a second process, the axon, which actually courses through the skull and synapses, communicates, in the first relay station of the brain, the olfactory bulb. And so we have a direct connection between the outside world and the brain by a single neuron.

Now as reductionists, we have reduced the problem of determining how the brain knows what the nose is smelling into two problems. The first problem is, How do we recognize this vast array of molecular structures that are defined as odors? Humans, for example, are thought to be able to recognize tens of thousands of discrete odors, and second, conceptually more difficult, is the problem of discrimination; that is, how do we tell, how do we discriminate, among this vast repertoire of odorous molecules?

The first problem, the problem of recognition, was at least solved at a superficial level by a fellow in the laboratory, Linda Buck. Linda Buck isolated the genes encoding the receptor molecules that bind odors on the tip of olfactory sensory neurons. And the identification of these genes provided significant insight into the problem of recognition. These genes appear to operate in a manner that's quite distinct from the receptor genes that operate in other sensory systems. So in the eye, for example—in this mélange of Magritte again—in the eye we are capable of detecting several hundred hues by only three photo receptors encoded by three genes that have an overlapping specificity for different wavelengths of light. In the case of taste, we have only about thirty genes. But in the case of olfaction, there are a thousand genes encoding receptors. The implication, then, is that the vast diversity of molecular structures that define[s] the universe, the repertoire of odorous molecules in our environment, cannot be accommodated, cannot be recognized by a small number of promiscuous genes, but rather we have a very large number of, so to speak, chaste genes in the chromosome. And this principle of a large number of genes accommodating odors threads through virtually all
eukaryotic species such that the simple worm has as many genes as does man. And so a genome which maximally may have within it 30,000 genes contains 3 to 5 percent of its genes in the form of odorant receptors.

Now the identification of odorant receptor genes, then, provides us a solution to the recognition problem. We recognize this diversity of molecules that are defined as odors by virtue of having in our genome a vast array of genes encoding odorant receptors. But the identification of the odorant-receptor genes provided us with a surprising insight in the more complex question as to how it is that the brain knows what the nose is smelling. For we could now translate this question into molecular terms, we could logically ask, How is it that the brain knows which receptors have been activated by a given odorant? For if we could come up with a cogent model by which the brain knew which of the thousand receptors were activated, we would have the beginnings of a discriminatory model.

The problem was simplified even further by an observation which demonstrated that within the sensory epithelium of the nose, each of the individual 10 million sensory cells, for example, mate only a single receptor gene. This allowed us to reduce the problem even further, for the problem of the brain discerning which receptors had been activated could now be reduced to a problem of the brain discerning which neurons had been activated. And by analogy with other sensory systems, we could perhaps argue that the brain could determine which neurons had been activated by segregated defined neurons in space, by creating a map in the brain.

A Cortical Map of Sensory Information

Now it’s been known for over a hundred years that the segregation of sensory modalities and submodalities is a basic principle of cortical organization in man. What we see here is that each of the individual sensory modalities projects to cortex in a discrete region. Not only do the individual sensory modalities project within a discrete region of cortex, but within a modality region, the somatosensory region, or the auditory region, there exists an internal map, and that spatial order serves to define both the position of a sensory stimulus in space and the quality of a sensory stimulus. If we return to the bat, for example, to the bat's auditory cortex, what we observe is that one important quality of auditory information, frequency, is linearly mapped along the bat's auditory cortex such that low-frequency sounds are on the right and high-frequency sounds are on the left. And so the brain uses the position of a signal to determine the quality of a sound bit. Importantly this map, this spatial map, is not proportional. You see within the center of this map a region encompassing about a third of the auditory cortex, which is finely tuned to sounds around 60 kilohertz. Sixty kilohertz is precisely the sound of the bat’s echo, so this disproportion in the allocation of space is uniquely designed over evolution to meet the specific ecologic and evolutionary needs of a specific organism.
This disproportion is also evident in the somatosensory cortex where . . . I cannot see this slide, can you? But in somatosensory cortex, what we observe is that there is a disproportionate representation of the body surface of the hand and face at the expense of the trunk and limbs. But most importantly the somatosensory cortex illustrates a second principle, and that is that this sensory map is not static, it changes with experience. So if one looks at the region of the sensory map in the brain of a violinist, his fifth digit on the left hand, which is used for fingering, occupies three times as much brain space as does the fifth digit in the normal population. And so in these sensory systems, we have a spatial map that meets both the evolutionary, ecologic, and experiential needs of the organism.

Now in the olfactory system the brain does not map in a traditional way the position of an olfactory stimulus in space. And relieved of this requirement, we asked whether the brain uses space to map the quality of an olfactory stimulus. And the answer is very clear: indeed it does. In these sorts of experiments, performed in the laboratory by Peter Mombaerts and Fan Wang, what we did was to genetically engineer, to alter a mouse such that all of the neurons that make a given odorant receptor and are therefore responsive to a given odor are turned blue. And they're turned blue both in the sensory sheet shown here—this is the internus of the mouse nose—but also along their projections as they course through the skull into the brain. And what we observe is that in the sensory epithelium, neurons that make a single one of the thousand receptors are randomly distributed, but order is restored in the brain. The processes of these neurons all project back and converge on a fixed point in the first relay station of the brain.

Importantly, that point is fixed, it's invariant, in all individuals in a species, and that point differs for all the thousand receptors expressed in the nose of an individual within that species. The consequence of this is indeed that there is an anatomic map. And it follows, then, that individual odors will activate a subset of receptors, which in turn will activate a subset of points in brain space such that the quality of an odor may be defined by unique spatial patterns of activity in the brain.

**Linking Sense to Activity**

But is that anatomic map . . . for after all I've shown you, is an anatomic map . . . is this anatomic map functional? Does an odor indeed elicit precise patterns of activity, and do these patterns of activity have any consequence for the behavior of an organism? To address these points, we turned our interest to an analysis of the representation of olfactory information in an insect brain. Insects are capable of rather complex olfactory-driven behavior that is mediated by a brain which is five orders of magnitude simpler numerically than the brain of a mammal. This is a mammal, and this is the insect *Drosophila melanogaster*, and what we observe is the nose of the insect is the antenna, and, remarkably, upon analyzing the genes and circuitry of the fly nose, we observe that despite the 600 million years of evolution that separate these two species, the basic principles of anatomy and
functional organization of the peripheral olfactory system appear to be shared between these two organisms.

Here we see the nose of a fly, and we’re lighting up cells that are expressed in a single neuron. And, indeed, what we observe is this very principle of the existence of multiple genes such that only one of the multiple genes is expressed in a single sensory neuron is preserved in the fly, and, moreover, all of the neurons that make a given odorant receptor each project back to a fixed point in the fly brain. So now we’re looking at the fly brain, and this is the functional and anatomic equivalent of the olfactory bulb that I described to you in the mammalian brain. And so despite the 600 million years of evolution that separate the two species, the two species appear to have evolved—in this instance, independently—the same basic solution to the organization of the peripheral olfactory system, suggesting that this solution is in fact one of a relatively few that solve this essential and rather complex problem.

And this organization indeed suggests that different odors in the fly will activate different loci—these are known as glomeruli—such that banana might activate one combination of loci which is overlapping but nonidentical, and the quality of an odor would be encoded by spatial patterns of activity. But is this anatomic map functional? And importantly, are there behavioral consequences to the activation of a discrete set of loci in the brain? To address this problem, we made use of the genetic facility of the fly, and in a series of experiments performed by Jing [Wang] and Allan Wong in the laboratory, what we were able to do is to express within all fly olfactory neurons a reporter protein, which fluoresced upon elevations in calcium in the neuron. And calcium is an indicator of electrical activity, and this allowed us to perform imaging experiments on the fly brain to actually examine the patterns of brain activity in response to odors in real time and actually look at what is happening in the fly brain in a two-photon microscope over real time with a sensitivity and a spatial resolution that is perhaps a thousand times greater than the sorts of human-brain-imaging experiments.

And so what Jing and Allan Wong were able to do was to develop a fly-brain population which allows this sort of imaging at the submicron level in an isolated brain preparation that functions for several hours under a microscope. And this is what it looks like, looking now at the antennal lobe of a fly in response to two odors, caproic acid, which you will see in red, and pyridine, which you will see in green. And indeed, what we see is that pyridine activates these two loci in the brain whereas caproate activates these two loci in the brain. So indeed there is a functional representation of the anatomic map in the fly brain such that different odors elicit different patterns of activity. The anatomic map is functional, but it is of behavioral consequence?

In a series of experiments that we recently performed with Seymour Benzer and David Anderson at Caltech, we began to study the observation that flies are averse—they are repulsed—by what is known as an alarm pheromone, an alarm
substance, that is emitted by stressed flies. We were able to demonstrate that a major component of the alarm substance is CO2. And we were able to demonstrate that, indeed, CO2 elicits upon imaging a very simple pattern of activity in the brain. It activates one specific locus, one specific glomerulus here, and through genetic chicanery what we were able to do is essentially inhibit activity in this locus and ask, If we inhibit activity within this locus what is the consequence to behavior? And the flies no longer show an aversive response to CO2. The inhibition of activity in all other regions of this antennal lobe have no effect upon the response to CO2. So indeed odors elicit distinct patterns of activity in the brain, and these spatial patterns are of behavioral consequence.

The Binding Problem

What I have clearly shown thus far is that different odors elicit different patterns of activity, and that these patterns of activity can be read out as specific behaviors. The implication is that the nervous system, in the case of olfaction, dissects and deconstructs the odor into its structural components such that a given odor is represented by multiple spatially invariant loci of activation. So in this model, apple may activate these three loci, [and] banana these three overlapping but nonidentical loci. And I can look down at this map of activity and with accuracy I can discern what it is that the fly is smelling. But I’ve accomplished this with my eyes and my brain. How does the fly do it? The fly has no eyes in his brain, so what or who in the fly brain is actually looking down upon this map and reconstructing this deconstructed image into a meaningful percept? This is a simple form of the binding problem, how bits of electrical activity are bound into a meaningful percept. And inherent in the binding problem is a related problem, the parsing problem. Consider these two odors. Each of these odors elicits an overlapping but nonidentical pattern of activity. If I expose an organism to a mix of two odors, that organism is often able to discriminate the two individual odors within the mix. But the pattern of activity that we observe in the mix now consists of five loci. How is it that the organism is able to segregate, to parse, these three points of activity and identify it as apple, and these three points of activity and identify it as banana, rather than looking down and seeing a new pattern of five bits of activity? This is the parsing problem.

Superimposed on these conundra is the added complexity that the ultimate percept not only reflects sensory input but brain context, experience, expectation, and even emotion.

Now the binding problem in its simplest form as I’ve described it is one that is shared by all sensory systems. Elegant early work by Hubel and Wiesel and independently by Semir Zeki and in talks tomorrow have shown us that in the visual system, to obtain knowledge of what it is seeing, the brain does not merely passively represent images reflected on the retina; rather, it must actively deconstruct and reconstruct the visual world. A visual world is deconstructed first in the retina and ultimately by parallel processing pathways that report the distinct
components of a visual image, thus color is represented in a region of the cortex known as V4, whereas cells in V5, as we'll hear tomorrow, are responsive to motion, and an adjoining region is responsive to form. V3 and V5 are indifferent to the color of a stimulus, and lesions on the color area V4 allow an individual to see an image with clarity, but only in shades of gray.

This segregation in the visual system immediately poses a problem not dissimilar to what I've described in the simpler olfactory system. It's a problem of reconstruction or binding. How is the spatial map read? How are bits of electrical activity integrated to allow for meaningful recognition of a sensory image? I've already argued that sensory input, the bottom-up process, is incomplete, that it often results in a meager and selective image of physical reality. The image is completed by the brain in a top-down process that brings experience and expectation to the binding process. And if this is true, then perception is, as originally suggested by Richard Gregory, only a hypothesis, a best guess that only asymptotically approaches reality. Our assumptions are based in part on input and in part from the brain's stored record, a record nourished by our visual experience.

Consider this particularly clear example of binding. Some of us will bind well and see immediately, and others will require more time to see this Dalmatian sniffing the ground in front of a tree. Once having seen, bound, and parsed this image, binding on second view will be instantaneous and you will never forget it. Experience shapes the way you integrate incoming sensory information.

Now if perception is really just an assumption, a best guess, then you can get it wrong. The Kaniza triangle is a classic example of what I call illusory binding. This is an image that consists of three black Pacmen, but this is not at all what you see. What you see, what you are focusing on, is an equilateral triangle whose sides should be quite clear to you. But these sides do not exist. Your brain is trying to use its preconceived notions of the visual world and makes what is not. This is an illusion.

Finally if perception is indeed a hypothesis it can be challenged, as Zeki has pointed out, and this is precisely what René Magritte did. Magritte defies the common sense of our brain deliberately and with great success. The painting Carte Blanche is confounding. It goes against everything the brain has ever seen, learned, or stored in its memory. We can have no preconceived notion here because the brain has no representation of this bizarre scene. It is an act of the imagination that fascinates precisely because we cannot find a solution. It is a classic example of nonbinding.

Let me then return to the biology of the binding problem. How is the deconstructed map in the olfactory system reconstructed? One disturbingly seductive model argues that the combination of signals from the antennal lobe might be brought together to report to a locus in high brain, so this combination of signals might connect to a single locus in high brain that would then provide us with a refined
olfactory image, the notion of a jasmine cell in olfaction analogous to a grandmother cell in vision. Indeed recent data in our laboratory suggest that the next level of olfactory processing does precisely this, that active loci in the antennal lobe, which are insular and segregated and have minimal communication, actually project to high brain to form a second map. But this second map is of a different character. The termini of neurons from the antennal lobe are no longer insular or segregated, but now interdigitate, and this sort of dispersive interdigitation affords the opportunity for integration, so it allows for the communication of these bits of activity in high brain and perhaps for their reading.

So these observations leave us closer to a solution, but we're left with a higher-order problem. For there is not likely to be a single master area to which all signals ultimately report in any sensory system. Moreover if there were, who would look at it, who would read its spatial image? Rather, the sensory representation, I would argue, is likely to be distributive, and how this distributive ensemble is read and ultimately elicits appropriate behavioral or cognitive responses is what Vernon Mountcastle has described as "the big in-between, the ghost in the machine." And it would be presumptuous of me, a geneticist, to try and approach this old and complex problem of the ghost in the machine. Who reads the image, who listens to the music? And so I will leave this task to the ghost busters, to Bill Newsome, John Searle, Christof Koch, who will address the problem of consciousness tomorrow.

Thank you.

**Question and Answer**

I'm happy to take questions.

[Question inaudible.]

The question is there are some odorants that have powerful effects on the organism. Have I examined those odors and how they map? Indeed it's true that there are a set of odors within each species that elicit innate and strong responses. And so pheromones in most mammals below primates, for instance, elicit an innate mating response. And indeed we have looked at pheromones, as you might imagine. Pheromones activate a set of neurons in a different nose, what I call the erotic nose, the vomeronasal organ, which projects to a different part of the brain. The main nose that detects the universe of odors projects to what you might consider to be cognitive brain, whereas the pheromones are detected by a nose that projects to amygdala, the emotive brain.
Introduction by Gerald Fischbach

Gerald Fischbach: Thank you, Richard. Just beautiful. Performing that link between molecules and mind is what our goal is in creating a new neuroscience institute in many different areas.

I first met Eric Kandel in the basement of a psychiatric hospital in Boston forty years ago. He's lost a pound or two since then but he hasn't changed one bit in those forty years. He was enthusiastic, full of ideas, creative and full of fun then, and he still is now. I have a feeling that Eric is just beginning to take off in his career and that the best is yet to come.

His early studies, what he was doing that day when I met him at [what was] affectionately known as the Boston Psycho was studying the neurons in a very simple nervous system with electrophysiological techniques. And his goal then was to create a simple alphabet of behavior using his training in psychiatry [and] very interested from the beginning in simple behaviors—in this case simple reflexes, the withdrawal of a siphon, a structure in *Aplysia californica* or habituation—and trying to determine how particular nerve cells in that simple organism's brain could account for that simple behavior. Now this, in addition to a simple nervous system, is a simple idea, but like many simple ideas this is an extremely powerful one. How does the function of individual nerve cells and simple neural circuits account for our behaviors? And it has had an enormous influence on the field over the years and on everyone in the field in this room. And from that simple alphabet, Eric has moved on to ask profound questions about learning and memory even in higher organisms.

He has won many awards, many honorary degrees, culminating, if you want to consider that, culminating in the Nobel Prize in the year 2000. Frankly, I believe the committee just ought to give everyone in this day-and-a-half symposium a Nobel Prize and get it over with, but Eric's was extraordinary. And you have the feeling that now from his position as a university professor he’s using that position well to expand his area of interest and influence in all directions.

He is collaborating I understand, or at least discussing, the nature of the brain and art with Jean Magnano Bollinger, who has been interested in this for a long period of time, with a recent exhibit of her scrolls. He's reaching into many different areas.
He is returning to his roots in psychiatry with an interest in schizophrenia, has been supported for a long time in this effort by the Liebers and the Lieber Center, and one has the feeling of a completion of aims and ambitions begun when he was a student in Boston in those early years, and now really with escalating energy and insight in his current lab work. He's going to talk to us today about his studies of neuroplasticity and memory. Eric?

**The Biology of Memory**

**Eric Kandel:** [inaudible] . . . that Denise is here to listen to you. She would not have believed me had I told it to her in private.

People ask me why I enjoy Columbia so much. I think listening to Gerry, listening to Richard, listening to Tom, you realize why neuroscience at Columbia is so spectacular. I feel that my own good fortune in this area owes a great deal to my wonderful friendships at Columbia, and you saw really superb examples of their work as they presented it.

So in my own talk this morning I'd like to outline aspects of our current understanding of the molecular biology of memory storage, and I'd particularly like to focus at a problem that's interested me recently, that is the persistence of memory. How come one can remember certain events, for example your first love experience, for the whole of your lifetime?

Let me begin by putting the biology of memory into a little bit of a perspective for you. As you know, learning is the process whereby we acquire new information about the world, and memory is the process whereby we retain that information over time. Most of the knowledge we have of the world we have learned, so that in good measure we are who we are because of what we learn and what we remember.

For biologists' interest in mind, like many of us, the study of learning has the further appeal in that it has broad cultural ramifications. It raises some of the vital issues that have traditionally confronted Western thought. What aspects of the organization of the mind are innate? How does the mind acquire new knowledge about the world? Serious thinkers of each generation have struggled with these questions and by the end of the seventeenth century two opposing views have emerged: The British empiricists such as John Locke argued the mind is a blank slate, it does not possess innate knowledge but that all knowledge derives from sensory experience and is therefore learned. By contrast, the German philosopher Emanuel Kant argued that the mind is born with a priori knowledge, preknowledge, that predisposes it to receive and interpret sensory experience in an innately-determined perceptual framework.

In the 250 years since the founding of Columbia University, it has become clear, even to our faculty, that the methods of philosophy could by themselves neither
distinguish nor reconcile these conflicting views because the issues they raise revolve around questions of what goes on in the brain when we learn. These questions require a direct examination of the brain and, as you've already learned and you will gather from the other talks at this symposium, in recent years neuroscience has done just that, and we now have a beginning understanding to some of these difficult questions.

To address these questions it is convenient to divide the study of memory storage into two parts, the systems problem of memory and the molecular problem of memory. In the systems problem of memory we ask where in the brain are the various memories stored, what different regions store different kinds of memory? In the molecular problem of memory we ask what are the molecular mechanisms whereby that storage occurs at each site? I'm going to focus primarily in the molecular problem of memory, which is the easier problem. The more difficult problem, as Richard alluded, is the systems problem, but I want to say just something about that.

**Stages in Explicit and Implicit Memory**

One of the major insights to emerge in cognitive neuroscience, in modern neuroscience, is the realization that memory is not a unitary faculty of mind, but exists in at least two major forms called explicit and declarative and implicit and procedural. Explicit memory is what you normally think of as memory, it's the recall of facts and events, it concerns itself with information about people, objects and places, it involves a particular set of structures in the brain, the medial temporal lobes, and a region deep to it called the hippocampus. And the defining feature of explicit memory storage is that it requires conscious attention for recall. This system is in parallel with a very, very different system, a set of storage mechanisms that are concerned with perceptual and motor skills, also involve simple forms of learning, such as classical conditioning, operant conditioning, sensitization, habituation. They involve a different set of deep nuclei in the brain, the amygdala, the cerebellum—the amygdala for learned fear, cerebellum for motor learning—and in the simplest cases of invertebrates the reflex pathways themselves. This is defined by not requiring conscious attention, and if you think about it much of your mental life is carried out without conscious awareness.

The existence of these two very different memory systems, which store different kinds of information, involve different kinds of brain structures and have a different logic, conscious recall and unconscious recall, raise the question, To what degree do they share features in common? Can molecular biology, with its ability to reveal homology relationships, which you learned about in the earlier lectures, delineate commonalities in these two radically different kinds of memory systems?

Now an initial clue to the fact that there might be features in common was actually suggested by William James when he pointed out that all memories have stages, a short-term memory, which last minutes to at most hours, and a long-term memory,
which lasts days and weeks. And we now realize that every form of implicit and explicit memory has these stages. Moreover, we know how to convert, in most cases, short-term to long-term memory. It involves repetition, practice makes perfect, just as your mother taught you. And three, we know that in both implicit and explicit memory storage, long-term memory storage differs fundamentally from short-term memory in requiring in its initial steps the synthesis of new proteins. And indicated here by indicating ribosomes, the machinery that is involved in protein synthesis, a point I'm going to return to later on.

This feature, the requirement for protein synthesis, is extremely conserved. You find it not only in explicit and implicit memory storage, but you find it in implicit memory storage in simple animals as well as complex animals. And that conservation suggests the possibility that if the requirement is so conserved perhaps the specific proteins are conserved. And if that's so, then the delineation of proteins in any single context might reveal the schema whereby short-term is converted to long-term memory. And if one can do it in several different contexts, one might be able to get a general idea of a principle of a mental process of how you convert a short-term to long-term memory.

And I would like to consider that with you by using two examples, one the marine snail *Aplysia* to consider a very simple form of implicit memory storage, that of a learned fear, and the other spatial memory in the mouse, the opposite end of the spectrum, if you will, an extremely complex behavior.

**Sensitization and Memory in *Aplysia***

Let me begin with studies of *Aplysia*. This is the marine snail *Aplysia*, this is not only a very beautiful snail but you can tell at a glance a highly intelligent animal. This is the sort of animal any one of you would select for the cellular study of learning and memory. The reason it is so attractive is not only because of its physical exterior but because its brain is remarkably simple. Your brain and mine is made up of a million nerve cells and, as Tom pointed out to you, these are interconnected in a set of complex ways. By contrast, the nervous system of invertebrates such as *Aplysia* consists of 20,000 nerve cells. The nerve cells are collected in clusters called ganglia, and each ganglion contains about 2,000 nerve cells. Moreover, a single ganglion controls not a single behavior but controls a set of behaviors, so the number of cells committed to a single behavioral act can be quite small, and the behavior I'm going to describe to you involves less than a hundred nerve cells.

Not only are there few cells involved in generating the behavior, but for reasons that one doesn't quite understand in the marine snail *Aplysia*, one encounters the largest nerve cells in the animal kingdom. These are gigantic nerve cells. Before I developed presbyopia I could see them with my naked eye. The largest cells are a millimeter in diameter. Because they are so large after a while you can recognize them and you can give them different names, Gerry, Tom, Richard, you can go
down the line—David Cohen—and you can return to the same cell in every animal of the species. So you can look at the same cell in a naive animal, an animal that's trained, and see exactly what has altered in the brain.

Using these advantages we focused in this simple animal with a simple nervous system on the simplest behavior that the animal has, a simple withdrawal reflex like the withdrawal of a hand from a hot object. The animal has an external respiratory organ called a gill, which is covered by a sheet of skin called the mantle shelf, which ends in a fleshy spout called the siphon. If you apply a tactile stimulus to the siphon you get a brisk withdrawal of both the siphon and the gill. This simple behavior, it turns out, can be modified by several forms of learning. And one of the principles that came out of it is even the most elementary behaviors are modifiable by experience. And with each form of learning there's a short-term memory, which does not require protein synthesis, and a long-term memory, which does.

I'm going to focus on a particular kind of learning called sensitization, a form of learned fear, in which an animal learns about the properties of an aversive stimulus and learns to enhance its reflex responses. So if you give the animal a weak tactile stimulus of the siphon you get a modest withdrawal of the gill, but you now give the animal a shock to the tail, the animal recognizes this as being offensive and it enhances its reflex responses in preparation for escape. So the same weak tactile stimulus that previously produced a moderate stimulus, after a noxious stimulus of the tail that same weak stimulus will now produce a much more powerful withdrawal. And the animal will remember that offensive event as a function of number of repetitions. If you give one stimulus you have a short-term memory, which lasts minutes, doesn't require new protein synthesis. If you give five trainings or more, you produce a long-term memory that lasts anywhere from days to weeks, and this requires new protein synthesis. So clearly we want to understand how do you set up the short-term memory and how do you convert it to long-term memory?

The first thing we did was to work out the neural circuit of the reflex, and I show you in simplified form what is already a very simple neural circuit. There are 24 sensory neurons that pick up from the siphon skin, they make direct connections to six identifiable motor neurons, and indirect connections to those motor neurons through inhibitory and excitatory interneurons. When we examined this neural circuit in some detail we were struck, as Tom first pointed out, by the invariance of these connections. There were always 24 sensory neurons, there were always six motor neurons specifically identifiable, and certain sensory neurons always connected to certain motor neuron and to certain interneurons. We saw here in reductionist form Kantians' view of the world. We see that built into the architecture of the brain is preknowledge, the neural knowledge for basic behavior.

But in turn, this raised the paradox, how do you reconcile this with the fact that this behavior can be modified? In order to see what happens with the modification, we looked dynamically on line, in time, what happens in the nervous system when the
animal learned sensitization. And we found that when you stimulate the tail you activate modulatory neurons of which the most important component is serotonergic, a transmitter also involved in your brain, and that modulatory system acts in the sensory neurons, including the presynaptic terminal, to strengthen the synaptic connections between the sensory neurons and the motor neurons. If you stimulate the tail only once there's a functioning strengthening, which persists for minutes. This doesn't require new protein synthesis. But if you stimulate repeatedly you release more serotonin and that activates genes in the sensory neurons, which ultimately give rise to the growth of new synaptic connections. This step requires new protein synthesis, the turning on of genes gives rise to proteins that are essential for the growth of new synaptic connections. So we see here Locke's contribution to the thinking of the gill-withdrawal reflex, and a reconciliation in a radically reductionist form of these two points of views; that is, built into the brain is the capability for neural action. But what is not specified in the genetic and developmental program is the exact strength of these synaptic connections and what environmental contingencies—such as learning—play upon is the ability to modify strengths. And you can do that with different forms of learning in both directions.

**Molecular Mechanisms of Implicit-Memory Storage**

So what is the molecular underpinning of this? In order to consider that with you, let me show you a blowup of the connections between the sensory neurons and the motor neurons. This is the sensory neuron. This is the motor neuron. Tail stimuli activate serotonergic connections. These serotonergic connections act on a receptor in the sensory neurons to engage a system of intracellular signaling mediated in this particular case by a second-messenger system called the cyclic-AMP system. That activates an enzyme in the cell—this is a way of carrying information from the cell membrane into the cell—it activates the cyclic AMP-dependent protein kinase, an enzyme in the cell, and that acts in the presynaptic terminal to strengthen the connections by releasing more chemical transmitter in the way that Gerry Fischbach described it for you. This does not require new protein synthesis, doesn't engage the nucleus of genes, this is a transient change, which lasts for minutes.

But what happens if you stimulate repeatedly? If you stimulate repeatedly the level of cyclic AMP goes up more, and this enzyme, the cyclic AMP-dependent protein kinase, recruits another friend, another enzyme, and they both translocate into the nucleus in order to activate a transcription factor, a control of gene expression, of the kind that Tom Jessell described for you. That transcription factor acts on genes that ultimately give rise to proteins that are responsible for growth of new synaptic connections.

So there are two points here that I want to emphasize for you: One is that learning recreates a program that Jessell described for you in development, whereby signaling pathways activate genes in order to give cells a sense of identity. Here in
learning we’re seeing how the outside world acts on signaling mechanisms to activate genes that give you a change in state of a preexisting neuron. So one thinks of genes as being the controllers of behavior. It's also important to realize that they're also the servants of the environment, they respond to external stimuli, and learning experiences produce long-term changes by altering gene expression. So insofar as you remember anything in these lectures—you probably want to forget my lecture—but insofar as you remember Richard's or Tom's or the subsequent lectures it's because gene expression is being altered in your brain. This is not heritable, you don't have to worry, your kids are not going to be contaminated by this, but genes are going to be altered in their expression in the brain. And the reason that alteration is important is because it gives rise to the growth of new synaptic connections.

And this is such an interesting point I just want to elaborate it for you. This is what Craig Bailey did in which he labeled individual sensory and motor neurons before and after learning. And you can see in *Aplysia* where the change is particularly robust that you see an outgrowth of processes in both the sensory neuron and the motor neuron. If you actually count the number of synaptic connections you see it doubling in the number of synaptic connections. So this is really quite profound because, to elaborate on my metaphor, insofar as you remember anything in these talks it is because altered gene expression gives rise to anatomical changes in your brain, so you’re going to walk out of this symposium with a somewhat different head than you walked into this symposium with, and all this without taking any drugs. Think of this, think of this, every single person in this room has a somewhat different brain than every other person, if only because of the environmental experiences. Identical twins with identical genes will have different brains because they've been exposed to somewhat different learning experiences.

**Explicit Spatial Memory in Mice**

So, so far I've talked about the simplest form of memory storage, implicit memory storage. What about explicit memory storage? How does it work and does it use any of this molecular machinery? Is growth also involved? Now simply to remind you, one of the reasons our fondest memories are recruited through explicit memory storage is because it recruits conscious awareness and it allows for mental time traveling. You know, I can sit back and I can think, forgetting about space and time, what it was like to be a kid in Vienna. I can remember coming to the United States. Each of you can sit back and recall events that occurred many, many years ago. You can remember specific places that you were and you can think of exactly what it felt like to be it, it's a remarkable experience how you can really overcome geographical and temporal barriers in order to reach back in your mind to these early events.

And I want to focus in a particular example of that, spatial memory, how one recalls space. A perfect example of mental time travel is to ask a London cab driver. I would not recommend you do this experiment with a New York cab driver. London
cab drivers know how to get around the city, and if you ask a London cab driver just to think in his head, or her head, how to get from Hyde Park in the south to Primrose Hill in the north of London, they will close their eyes and if you image them it will light up their right hippocampus, the region that is involved in spatial memory. Moreover, and this will come as no surprise to a sophisticated audience like you, if you ask experienced cab drivers compared to relatively new cab drivers and image their hippocampus you will find that experienced cab drivers will have a larger right hippocampus than will inexperienced drivers. So even in people there is an enlargement of the structure with continued use.

We can explore this in experimental animals. The mouse has a perfectly good spatial memory system, in fact it's exceptionally good considering it's a mouse, and it has a beautiful hippocampus, which is a minor structure of your own. So the animal can find, for example, one hole out of forty that leads to an escape hatch. And it finds it very readily. And we can really begin to try to understand how information about space gets into the hippocampus. And this is really quite interesting as it follows naturally from what Richard told you before. There is for each classical sense modality, for touch, for vision, for olfaction, a topographical, a map-like recreation in the brain, an internal representation that is organized so that neighborhood relationships are preserved. But space is a fiction, space is not a single modality, space is a composite that is put together only in the head so you can reconstruct it. And it requires a number of different senses, visual, tactile, position senses, as well as olfactory. This is put together in a series of cortices, and finally projected into the hippocampus where the map is combined in the most coherent way in a particular subregion called the CA1 region. You can record from single cells in the hippocampus, as John O'Keefe first did in 1971, and see how the spatial map develops. And it's really quite remarkable.

This is a replication that we did of the O'Keefe experiment he first carried out in rats. We worked on mice because we can do genetic manipulations, but this is essentially what O'Keefe found. You can image an animal, a mouse or a rat, moving around in an enclosure with specific markings so it orients those markings, and you can record from a number of different nerve cells simultaneously in the hippocampus. And as the mouse moves around you will find that different cells fire when the animal assumes different positions in space, so some cells will fire here, bur-bur-bur-bur-bur-bur, others will fire when the animal is here, bur-bur-bur-bur-bur-bur-bur, others will fire when the animal is here, bur-bur-bur-bur-bur-bur, and if you bur-bur-bur-bur-bur-bur-bur enough you will be able to see that every single position in space is occupied by one cell that is responsible for that. And if you record from a hundred cells, as Matt Wilson and others have done, you can predict where the animal is in space from its firing pattern.

And I just show you three examples. The yellow pseudocolor means that the animal is moving around in this space but the cell does not fire, so cell number one fires when the animal is at six o'clock, cell number two fires when the animal is at nine o'clock, and cell number three fires when the animal is at twelve o'clock and
somewhere in the center. So this is very nice. What is fascinating about this is that unlike other sense modalities every time you move in a space you have to learn the space de novo. And, in fact, you find if you put the animal in an enclosure, within 15 minutes it forms this internal representation, and in optimal circumstances, and I will define for you in a moment what optimal means, that animal will be able to retain that map over a long period of time. You take it out of the space, keep it out of there for several weeks, bring it back into the initial space, it'll replay exactly that same internal representation. You can take it out of the space in which it's familiar, put it into a new space, it'll form a new map with some of the same cells and some different cells, put it back into the original space it'll replay the original map. So this is quite interesting, really a learning process that's going on here and it's distinctive for each space.

Attention and Spatial-Memory Stability

And now we can ask, How does attention fit into all this? Is attention important in the formation of the map, or is attention important for the stabilization of the map in a long time? So it is important for first forming it, or is it important for maintaining it? In order to explore that ability Cliff Kentros developed a graded series of attentional demanding tasks. The first was one just basal attention, the animal walks around without your doing anything particular in order to draw its attention. In my case it would be like my walking around in a fog, as I usually walk around. The second thing is you throw food pellets into the enclosure so the animal's attention is aroused a little bit. A third degree of attention is you in addition to having the animal move in its usual space you introduce a discriminating space, and finally you can really draw the animal's attention by having it do a spatial task. And that is the animal walks around in its enclosure, all of a sudden noises and lights come on. The animal hates that, mice are not like neurobiologists, they don't want the attention, they don't want the publicity, they want to get away from that, and the only way they can do that is to sit onto an unmarked goal region and sit on it for a couple of minutes. That turns all these stimuli off.

And now we ask, How does attention affect the map? And we found that irrespective of the degree of attention, the map always forms and is stable in the short-term. So even basal attention is sufficient to give you a formation of the map and stability in the short-term. But what attention is required for is the long-term stability of the map. And I can illustrate this with one example: So these are two cells taken from the two extremes that I showed you before, from the basal case and the optimal spatial-attention task. If you take the animal that is just basal attention, just barely paying attention, you can see that this cell fires in a particular position, you take it out, three hours later it fires in very much the same position. But in the absence of attention, if you now look at the same cell the next day, the next day, the next day, you see that it shifts every day in firing a different position. In contrast, if the animal is paying optimal attention, not only does the map form and is stable in the short-term, but it's like a cookie cutter, every day it fires in exactly the same position. And if you look at the systematic
relationship you see that the long-term stability of the map absolutely requires that
the animal pays attention, that the map is not stable, as I will show you later on, the
animal cannot, without a stable map, remember spatial locations in the long-term,
but if the map is stable it remembers the locations in the long-term extremely well.

**Dopamine as a Candidate Mediator of Attention**

But that of course raises the question, Can we use this as a radically reductionist
approach to studying attention? How does attention work, how does it play itself
out on the neural system, how does it work on the cells in the CA1 region of the
hippocampus? And to do that one can take advantage of the fact that we actually
know a modest amount of the system's properties of the hippocampus, how
association cortex feeds into the hippocampal formation, and how it feeds into this
particular region. And one of the characteristic things we know about this region is
that this region is extremely important for spatial memory. Larry Squire has shown
that if patients have a lesion restricted to this area they have a profound deficit in
explicit memory storage, including spatial memory. Moreover, the pathway into the
hippocampus, the Schaffer collateral pathway into the CA1 region has been
extensively studied, and it shows a characteristic alteration in synaptic strength not
too dissimilar [to] what I showed you in *Aplysia*. This is called long-term
potentiation [LTP], and we know a lot about it, and we know that if you interfere
with that you interfere with spatial memory.

Now a large number of labs have contributed to working out the signaling
transduction pathway for LTP in these hippocampal neurons. And in brief it begins
with calcium inflow into the postsynaptic cell, which engages again a kinase that
gives you a transient strengthening of synaptic connections. The details differ from
implicit memory storage but the principle is very much the same, a functional
change in synaptic connections. But if you activate it repeatedly the calcium
activates this enzyme that synthesizes cyclic AMP, you recruit the cyclic AMP-
dependent protein kinase, you activate genes—actually many more than I indicate
here, this is a simplification—and it gives rise to the growth of new synaptic
connections. What is also fascinating is that an absolute requirement for turning on
genes is the conjoined action of a modulatory input, analogous to the serotonergic
one, a dopaminergic input that is required to turn on gene expression.

So we can ask the question my gosh, this looks awfully similar to the salient signal
mediated by serotonin. Is this how attention mediates part of its action? Does it
come in through the dopaminergic input to the hippocampus? So Cliff Kentros
asked this question, he took animals in a completely nonattentive mode and gave
them a drug that stimulates these receptors specifically, and he was able to
improve this attention deficit disorder in the mouse. Moreover, if you took animals
that paid slightly more attention and compromised this receptor with an inhibitor of
the receptor, you could compromise the stability of the place cell even more. So
here you could increase the stability of the place cells, and here you could
decrease the stability of the place cells.
And now we could do something very nice, we could say, "Look, let's use the power of genetics in order to explore this, let's affect one of these steps and see how this affects the spatial map." And also, since we can use the animals for multiple purposes, we have lots of these mice, we can see how it affects spatial memory storage. And what Ted Abel did when he was in the lab was to produce a line of mice that were compromised in the cyclic AMP-dependent protein kinase, so the whole later steps, all the later steps involving the turning on of the genes and the growth of synaptic connections, could not occur. If you looked at the spatial map in those cells in those animals you found that they formed perfectly well within one hour, because that does not require the dopaminergic input and attention, but it was unstable at 24 hours. If you now in these same animals asked, "To what degree is the spatial map necessary for spatial memory?" you find that spatial memory also is impaired. Perfectly good in the short-term when the map is stable but compromised in the long-term.

So this is really quite interesting because it suggests that despite the fact that explicit memory storage and implicit memory storage are radically different in terms of the neural systems that use them, and the nature of those neural systems, the storage mechanism per se shares core features in common. In each case a signaling system that involves importantly the cyclic AMP-dependent protein kinase activates genes, they give rise to the growth of new synaptic connections, and a modulatory system is importantly recruited to trigger the long-term process.

What is interesting and really sort of struck us as we thought about it is we even saw a fundamental difference between implicit and explicit memory. Think of it, implicit memory, like explicit memory, uses a modulatory system in an important fashion. But we would argue that the difference between the two is how that modulatory system is recruited. In implicit memory storage it's recruited in a bottom-up fashion, unconsciously, if you will, by activating the tail, the tail sensor neurons contact directly the serotonergic cells. By contrast, in explicit memory storage, we know that the information comes from the cortex itself, from the prefrontal and posterior parietal cortex, then projects down in the dopaminergic system, so you have a top-down influence. So the difference between implicit and explicit is between unconscious attention, if you will, and conscious attention.

**Local Protein Synthesis and New Synaptic Growth**

But given the fact that in various forms of learning, and this sort of schema has now been shown to apply to a number of different learning processes, both implicit and explicit, it really raises a fascinating question that I want you to think about. If long-term memory involves gene expression and therefore the nucleus, an organelle that is in principle in contact with every synapse of a neuron, does that mean every time you turn on a long-term process in your brain it must necessarily involve every single synapse, that it's a neuron-wide process? If that was so it would limit tremendously the computational power of the brain. A single neuron
has not one but a thousand different synapses, which contact a number of different target cells. Does that mean every time you throw a switch for the long-term process you throw the switch for all of these, or can you use a transcriptional mechanism and restrict expression to some synaptic terminals and not others?

This is a question Kelsey Martin addressed when she was in the lab. She reconstructed, really based on a methodology that Gerry Fischbach developed, the gill-withdrawal reflex in dissociated cell culture. She took a single sensory neuron, two motor neurons, quite distant apart, the critical elements of the neural circuit of the gill-withdrawal reflex in *Aplysia*, and puffed on serotonin and showed she could simulate the learning process perfectly well in the dish. When she stimulated once briefly she simulated the short-term facilitation, lasted minutes, restricted to this synapse, nothing there. More surprisingly, when she gave five pulses she produced a long-term facilitation that lasted days. This required new protein synthesis. This required alterations to the gene expression. If you block CREB in the nucleus you didn't see this; moreover, it involved the growth of new synaptic connections. But if you looked here you saw nothing. So you can get a transcriptionally-dependent long-term process with the growth on new synaptic connections at one set of terminals and not others.

And that raised the question, How does this come about? And Kelsey figured out that a signal goes back to the nucleus to activate gene transcription. Gene products are sent to all the terminals, but only those terminals that are marked by the short-term process can utilize those gene products productively, those proteins and messenger RNAs that are coming down productively, in order to grow new synaptic connections.

So that raised the question to the next level: What is the nature of the marking signal? And Kelsey made very good progress in working even this step out. She showed that when you activate a set of synapses with serotonin you send a message back to the nucleus, gene products are sent to all processes but only those that are marked can utilize that successfully for the growth of new synaptic connections. And the mark has two components to it: A cyclic AMP-dependent protein kinase, the same kinase that I showed you before, is necessary for the growth of new synaptic connections. But there's another surprising component. We've known for many years that there is, in addition to the cell body, there is in each synapse a machinery for locally synthesizing proteins. And her study made it clear that this machinery is part of the mark, that if this is not activated this growth occurs but is not maintained. So even if this growth occurs by just marking the cyclic AMP-dependent protein kinase, if you don't allow protein synthesis to occur that growth retracts in front of your eyes. So if you block protein synthesis you will see that within a day or two that synaptic growth retracts. Now this is a remarkable result because it made one realize for the first time that setting up the long-term process, growing the next synaptic connections, is only part of the machinery for maintaining something through the lifetime of an organism. You need a local protein synthesis in order to maintain that process, to perpetuate it over time.
A Prion-Like Mechanism to Carry Memory

And that of course raised the final question: How is this perpetuation achieved? And this is where Kausik Si entered the picture a couple of years ago. He found that what maintains the local protein synthesis is a protein that’s a regulator of local protein synthesis, it’s like a transcriptional factor is to genes, this is to protein synthesis. It has a funny name this protein, it’s called the cytoplasmic polyadenylation element-binding protein. He found that there was a new form of this protein in the brain, and this was absolutely essential for activating the protein synthesis necessary to do this, and if you blocked it you grew connections but they were not maintained. That was very nice, it gave one an insight as to what regulates protein synthesis, but then he asked himself the question, What does this have to do with maintenance, how do you get from here to maintenance? And he looked very carefully in the protein and he found amazingly that this protein looks like proteins that are involved in prion formation. Now prions, as you probably know, are horrible things. They cause mad cow disease, they cause Jakob-Creutzfeldt disease in people, they are proteins that can maintain themselves through self-perpetuation. And the reason they maintain themselves is because they cause death and destruction.

Why would you want to find something like this working in the brain? Well he found that it works in the brain in a very different way, it works to do good, it uses the same mechanism but for beneficial purposes. But first let me tell you something about how prions work and then I will show you the novel variation that’s been found in this context.

Prions are distinctive—and this is what Stan Prusiner first pointed out—for existing in two conformations, A and B. They could convert from one to the other, they’re interconvertible, and one form, and I indicate this in yellow, is dominant, it can act on the recessive form and cause the recessive form to convert into a dominant form. And the dominant form tends to form little aggregates. Now this is interesting because it turns out that in this context, and we have reason that something similar might be occurring in the mouse as well, the dominant form is in fact the normal functioning form of the protein. The other form is the precursor form. So the way we think of it now is indicated in this diagram. You give five puffs of serotonin and you set up the long-term process, but only in the synapses that have been marked—and I'm going to focus on this one—can you utilize the gene products produced by the expression of these genes effectively. And the way that works is that a marking pulse causes the synthesis of this protein, CPEB, and most of the copies of that protein are produced in an inactive precursor conformation. But either stochastically or with the help of chaperone proteins conformation A is converted to conformation B, and serotonin is critical for that. And once it’s converted to conformation B, which is the prion active form, the dominant form, it feeds back on these inactive forms and causes them to become functional proteins forming these small aggregates. And it is these aggregates that bind to messenger...
RNAs that leads to their modification so they can be translated and be used to give rise to proteins that stabilize the synaptic growth.

So when you think of it this is really a very nifty mechanism. You can change at will, if you will, certain selected synapses and neurons without touching others, and you can do that by using a perpetuation mechanism that in principle can carry this a very, very long period of time.

What is fascinating about this is the study of the prion in a completely different context, in the study of memory storage, has revealed a new class of properties about prions that was never previously anticipated. And that is that a physiological signal can regulate the conversion, and that the form that is the self-propagating form is not a killer form, but it is the good, healthy, functional form of the protein that allows the memory to be carried forward in time. And it suggests that the possibility that we may be describing here one part of a sort of larger class of proteins in which the self-propagating form of a prion mechanism is used to create a functional protein that can exist in its self-replicating form for long periods of time, and that might be involved in steps of development, in neuronal identity, in persistent gene regulation, and in a number of other contexts.

So one thing you obviously want to ask yourself is, this is the way it is utilized in this one example in implicit memory storage is *Aplysia*. Does it also apply to explicit memory storage? Is this how attention stabilizes the memory? And we obviously do not as yet know the answer to this. But we have some preliminary clues. Martin Theis has found that there is a homologue of this form including an amino acid sequence that looks like it might have prion capabilities that is present in the mouse brain. It is present in the hippocampus in particular. Moreover, we have found that dopamine alters the level of expression in the mouse brain—Martin has found this. So this suggests the very interesting possibility. I only give it to you as a suggestion because the evidence for it is extremely weak at this point, that dopamine, again working like serotonin, not only mediates an attentional mechanism but this attentional mechanism might be able to stabilize place fields by perhaps recruiting a mechanism very similar to one I described to you in *Aplysia* in which a prion-like mechanism acts to stabilize synaptic connections in the brain and allows you to remember spatial relationships for the long-term.

**Conclusion**

So I've described to you how molecular biology—and molecular biology, which I really learned from the hands of these two masters that you saw in front of you—has really been able to take a complex mental problem and in reductionist systems use it in order to see the generalities of biological problems, a theme that you've heard repeatedly this morning and will hear again this afternoon.

What is difficult in memory storage is to work out the systems problems, which Tom and Richard began to address. And this is one of the reasons we're very
pleased about the Kavli Institute, which is going to be concerned with the systems problems of neurobiology. We want to know in detail, How do the various sense modalities actually come in and combine in the hippocampus to give rise a sense of space? This is an enormous challenge and we’re just at the foothills of what is a great mountain range.

So let me simply conclude by pointing out the colleagues in my lab: Naveen and Cliff were involved in the spatial map studies; Kelsey, Maurizio, and Amit Etkin were involved in the synapse specific facilitation; Martin Theis was involved in the recent work on the mouse; Kausik Si single-handedly opened up this self-propagation CPEB story; and, we were fortunate to have the collaboration of Bob Muller, Craig Bailey, and Susan Lindquist. Susan was involved in the prion story. And I've been very fortunate to have the long-term support of the Howard Hughes Medical Institute and recently Columbia's been privileged to receive the Kavli Institute.

Thank you very much.

**Question and Answer**

I'd be delighted, if you are not exhausted and starved, to answer any questions that you have.

[Question inaudible.]

You have both the body and the . . . first of all, neglect is a posterior parietal cortical problem, it is not a hippocampal problem. And you really have an external image. You know the classic experiment of standing at the steps of the cathedral in Milano in which the patient can recognize all the landmarks on one side, can't recognize it on the other side. You now ask him to stand in the other side of the piazza, he reverses what he does. So he's got the knowledge in his head, he just cannot handle the spatial relationships.
Brain and Mind
May 13, 2004
Richard Mayeux, MD
Introduction to Session II

Four Vignettes

Richard Mayeux: Good afternoon. Welcome to the afternoon's session. My name is Richard Mayeux. I'm going to be your moderator and host.

Just as I was coming up here I was handed a memo. There seems to be an electrical problem with the lights, and if you have a cell phone or a beeper, and it happens to go off you'll be electrocuted. So I'm going to ask everyone to turn your cell phones and beepers off. Thank you very much.

This afternoon's session will deal with brain function and disease. I'd like to start off by giving you a couple of vignettes.

Charlie had been an ambassador until he retired at age 76. He read biographies. His main complaint on the days that I visited him was that he had to reread several pages to remember where he was the night before. Every day he asked me if I had a cure for that problem, and he died about a year after that while sailing off the coast of Nantucket. He had Alzheimer's disease.

During her twenties, while working as a broker in a prestigious firm in New York City, Deborah began to experience periods of severe depression accompanied by a loss of interest in her family and her friends and in her work. She found it increasingly difficult to concentrate. She experienced a loss of energy, difficulty sleeping, and eventually lost that job. By age 30, these bouts of depression were interrupted by periods of excitement, irritability, insomnia, and rare hallucinations. And she's still out of work, although she's functioning somewhat better under treatment. Deborah and her husband Matt are now trying to decide whether to have a child. Deborah has bipolar or what's called manic depressive illness.

Michael's birth was very much anticipated by Janet and David. However, from an early age Michael's parents found him to avoid eye contact, to engage in meaningless, repetitive behaviors, didn't play with other children in the nursery, nor did he seem to have any desire to be contacted by anyone, even his parents. His speech was limited and he spent many hours focused on a particular piece of furniture in the house. Michael is now 6, he's in a special school, and he's been diagnosed with autistic disorder.
Joe is an outdoorsman. He built his home for his family with his brother, [and] he's a construction worker in Westchester County. He had a big passion for fishing and for extra-long camping trips in upstate New York. He and his family would often go for long walks in the Adirondacks. At age 68, though, he began to slow down, he was beginning to look older, stooped, he developed a postural tremor. Once a diagnosis was made and his treatment was started, he resumed many of these activities, although Joe was never quite the same person that defined him in those early years. Joe has Parkinson's disease.

The Burden of Brain Disease

These are just a couple of examples of the devastating effects that diseases of the nervous system have on the lives of people. The burden of brain disease—and by that I mean all neurologic, psychiatric, and neurosurgical diseases—is high, one in seven people have a disorder of the brain. According to the Global Burden of Disease 2000 study, 20 percent, a fifth, of the years of life lost by premature death or lives lived with disability in the population can be attributed to stroke, depression, alcohol abuse, and Alzheimer's disease. Notably, after heart disease, those are the top four, so in the top five, four of them are neurological disorders or psychiatric disorders.

Studies of most brain and nervous-system disorders, such as these in twins, suggest that at least 50 percent of the societal burden is likely to be genetically influenced. But most of these common neuropsychiatric disorders are genetically complex. They can present as a Mendelian disorder where they follow the predictable rules of genetics, or they can be non-Mendelian, in which the signs and symptoms occur in what seems like a sporadic fashion, although there is this tendency to aggregate in families. Variations in several genes can lead to the same overall disease manifestation, but there are also disorders in which different types of mutations in the same gene result in different or unique manifestations and sometimes even different diagnoses.

Over the past two decades we've learned a great deal about Alzheimer's disease, Parkinson's disease, ataxia, depression and schizophrenia from the study of genetic variation in families. For example, inherited variations in four genes cause Alzheimer's disease differing only by the age at which you develop the condition, ranging anywhere from 20 to 80. Similarly, variations in four genes cause Parkinson's disease. But both sets of abnormal genes affect a common metabolic pathway that leads to regional brain degeneration. Family studies currently in progress indicate that variations in several additional genes will be identified in the next year or so, and each is expected to contribute to the complex story.

Recent studies of families of depression and schizophrenia have also identified promising leads to abnormalities in the brain that contribute to the cause of these disorders. Early-childhood trauma may trigger depressive episodes in later life among people who have a particular variation in the serotonin-transporter gene. At
least four of the five susceptibility genes for schizophrenia appear to be involved in synaptic plasticity, signaling, and possibly glutamate neurotransmission.

Mutations in five different genes cause the most common form of hereditary neuropathy or nerve damage. It's called Charcot-Marie-Tooth. Duplication of a base pair or the presence of an extra segment of DNA in a gene called the Peripheral Myelin Protein, or PMP22, leads to the dominantly inherited form of the disease, which is most common and in adult years can be quite disabling. However, if you're missing a segment, if there's a deletion of the segment in the same gene, it leads to a disorder that's much less disabling called Hereditary Liability to Pressure Palsies. It was first discovered in tulip growers in the Netherlands, who after spending long hours on their knees working would develop a foot drop with subsequent complete or partial recovery. And on the other hand, if you delete a single base pair from that same gene, you call a devastating disease of childhood called Dejerine-Sotas, which is an inflammatory demyelinating neuropathy.

**Introducing Session Speakers**

The speakers for this afternoon's session on brain function and disease are going to take you through their own series of fascinating discoveries about the diseased brain. Using modern imaging techniques as a window into the unknown, Dr. Rapoport will discuss developmental alterations in children with hyperactivity and psychosis that are leading her team to a better understanding of the causes of these diseases.

Huda Zoghbi is going to discuss her work on Rett syndrome as a neuropsychiatric disorder in which a mutation on the X chromosome can alter gene expression, chromatin composition, and chromosomal architecture. Understanding this disorder will probably shed light on the fundamental properties of nerve cells, but also on the spectrum of autistic disorders. Continuing in this discussion will be Dr. Rutter. Autistic disorder, Asperger's syndrome, Rett syndrome, [and] pervasive developmental disorders represent a complex of developmental disabilities that occur early in life. Sir Michael Rutter will discuss his work on the genetic basis of these disorders and their pathogenesis.

Substance abuse is a massive public-health problem affecting individuals of all ages. Advances in brain imaging have allowed Nora Volkow to visualize and quantify the deleterious effects of drug abuse on the brain. Using this information she and others hope to develop strategies to prevent this deadly affliction.

More importantly, I think what you're going to hear this afternoon should give you hope, the hope that someday devastating diseases of the brain and nervous system will be detected earlier in life, perhaps even before birth, and effectively treated before the onset of disability, [and] the hope that our neurologic and psychiatric hospitals can become centers for prevention of disease rather than for
the treatment of chronic and progressive disorders. Progress, though, takes place in carefully planned steps. The scientists that will share fruits of their efforts with you this afternoon are making great strides and are well on their way to satisfying our hopes and dreams for a better quality of life, perhaps someday free of the most common disorders of the brain and nervous system.
Brain and Mind
May 13, 2004

Judith L. Rapoport, MD
Brain Development in Healthy, Hyperactive and Psychotic Children

Introduction by Richard Mayeux

Richard Mayeux: Our first speaker, Judith Rapoport, received her undergraduate degree in psychology from Swarthmore, her MD from Harvard, and was named chief of psychiatry at the National Institute of Mental Health in 1976, and she's been there since that time. She's also a clinical associate professor of pediatrics and psychiatry. Dr. Rapoport’s research interests include obsessive-compulsive disorder, childhood schizophrenia, diagnoses in child psychiatry, biological aspects of pediatric psychiatry, and pediatric psychopharmacology. Her most recent interest is in brain imaging of the development disorders of brain and language. She has written extensively on her findings [and] is on the editorial board of many scientific journals. The title of her talk today will be "Brain Development and Childhood Psychopathology." Dr. Rapoport.

A Top-Down Approach

Judith L. Rapoport: Thank you very much. It's a real pleasure to be here and honor to be on this wonderful program as well as a great pleasure because I have many very good friends on the faculty and in the audience.

What I want to share with you today is very much a top-down approach to psychopathology. We have many disadvantages from the clinical point of view. We're not certain about our clinical categories, [and] we have complex diseases, as have already been mentioned, where even when genes are found, they tend typically to be of small effect. And we have another disadvantage known to clinicians of the time-consuming nature of the studies that are involved. Nevertheless, it turns out that there is some enormous hope owing to some of the new technology that lets old questions that have become perhaps tired or at least frustrated become totally new and exciting. And this rejuvenation of the field and the sea change that's taking place among clinicians like myself is really what I'd most like to convey.

As I was saying, technology changes everything, as you can see.
I think most of us have had a long-held assumption because of converging evidence from neuropsychology and other neurological exams that the more severe child psychiatric disorders do in fact reflect subtle neurodevelopmental impairment that someday could be reflected in abnormal brain development. But the advent of brain MRI which doesn’t use any ionizing radiation and which most children tolerate without any adverse effects really permits the study of both normal and abnormal development, and it’s hard to describe the way this entry into the brain has expanded all of our horizons, our dialogue with many other disciplines.

And what I just want to show you is a series of studies on normal and abnormal development, restricting the studies, although we’ve done many others, to anatomic brain development. And what’s more, I’m not going to talk about the many other studies of cognitive development—clinical response, drug treatment, etcetera, family studies—that would accompany all of these because I want to just try to focus on one relatively simple theme.

The other thing that’s important to point out is that I’m afraid my other presenters are not off the hook in terms of talking about brain development because, as you can see from this diagram, there’s, starting with conception, an enormous number of different stages in the development of the brain from neurulation, neurogenesis, synaptogenesis, cell migration, programmed cell death, and so on. But what we did hope that we might see reflected was perhaps some sense—and this is inferring because I will only say once, but you should always assume I’m saying it—that we’re unaware from an MRI level of what the cellular basis is of what we change. But we are prone to model building, and we assume that some of these may reflect competitive elimination of synapses, effect on dendritic axonal arborization, and, of course, trophic glial and vascular changes that go along. But then I’m also pretty much only going to show this very thin, pale ribbon that you see at the right of our study, which actually goes down perhaps to four years in some cases and up to now 24 or 25 in others. So the other presenters who were hoping that I would present more about brain development are still . . . this is the only thing to reflect the very many and important processes that I’m not addressing.

MRI Studies of Normal Brain Development

What we started to do 15 years ago was to have a large series of subjects with different disorders in healthy children simply [to] have a brain MRI every two years, with a machine that fortunately turned out not to be a lemon—and some of them are—that has lasted with us through all of this time, and this is simply a diagram to give you a feeling of the healthy database which so expands every month that it’s already several months behind. But the pink lines are females and the blue lines are males, and if a dot is connected it means that this person has had a large number of rescans. And for each of our cohorts, we have some comparable kind of data, and the patients also had scans and rescans every two years.
And what made our life possible was at the same time a wonderful package for measurement of brain MRIs was being developed. There are now many competing packages, but at the time when we started, one of the most prominent—and the one we have stuck with for some of these studies—has been that from the Montreal Neurological Institute, where they combined two different approaches to measuring brains, one that simply at the top uses prior information from what's where in the brain, from neuroanatomy, and another one that simply classifies tissue of brain—gray matter, white matter, or CSF—and by marrying these two gets more information than you're likely to from either one alone. Most important to us, since we now have more than 5,000 scans across our various studies, it is fully automated, and their supercomputer can work away doing hundreds on a weekend while we're gainfully occupied doing something else. That's a very endearing characteristic of this system.

As you can see from the top left-hand corner, there's much to be gained from prospective studies. Your brain size varies more than your shoe size or your glove size, and as a result when you have brain data across samples, cross-sectional data—just comparing at any point in time a control group, for example, and a patient group—most of the studies, because of the noise within differences, are hopelessly underpowered to be able to make any statement about diagnostic differences, whereas when you start to include your longitudinal data and some singleton data, thanks to statistical advances what you can start to see are that these curves are in fact not straight. They're nonlinear, at least some of them are, and you start to see all sorts of interesting things.

This is parts of the brain, crude parts in this study, frontal gray, parietal gray, temporal gray matter, and several things come out here. First of all, that if you have a large number of subjects, you start to see that different parts of the brain mature at different times, with the temporal lobe maturing significantly later than the other parts. Interestingly—and these are also statistically significant—the pink reflects the females in the group, and as you can see, there's also some timing difference which for frontal and parietal are also statistically significant. A talk that I am not going to give now, but could very entertainingly and easily on some other occasion, are sex differences in brain development, which are very provocative and may well bear in part on I think some of the very interesting data that Dr. Rutter is likely to talk about. At any rate, the strength of having this normative data has been extremely important to us.

A lot of other spin-offs from this study have come out—for example, what's the relationship to intelligence. And the interesting part is that while IQ is related to the height of the slope—that is, how the volume of almost any structure . . . it has no effect on the shape. We had wonderful hypotheses that intelligence might relate to either the steepness or the ascending part of the steepness of . . . no such luck, it remains exactly parallel.
We speculate, of course, about what may relate to the differences underlying and
the molecular geneticists in the audience, I think, could come up with pages of lists
of candidate processes, signaling systems and so on, that could possibly be
involved. But I just want to mention that the pioneering work of Peter Huttenlocher
and his students . . . looking at the very scant number of postmortem brains across
childhood and adolescence showing both the rate and some regional differences
across different parts of the brain, both for the timing of development and, more
important, for the timing of the decrease with the overproduction of synapses to
healthy adult levels.

Another thing that we've learned—not shown on the diagram that I just showed
you—are the hordes of identical and fraternal twins, who are also going through
this process, so that we are in the process of studying the heritability of the slopes
for different parts of brain development at different times, which we expect will be
different, the heritability for different brain structures at different times during this
age period. But some things are clearly heavily genetic—for example, the corpus
callosum, some white matter structures, easily you can tell, even if pictures fall on
the floor, you can know which twin to put together easily from some structures. And
even with our limited amount of twin data, from this old slide, the heritability is fairly
high for most of the structures, even in childhood. So the main point so far is just to
tell you that prospective studies give a unique sensitivity which we've never had
before, that different brain regions develop at different times, and that there's some
regional difference in heritability.

MRI Studies of Child-Onset Schizophrenia

Now having had now my very short course in what we've learned about normal
brain development, I want to go on to talk about some of the disorders that we've
been tackling. I'll never do justice to the issues with the brain and schizophrenia
because almost every part of the brain that's been examined has proven to have
some kind of abnormality of this devastating disease that is so expensive to society
and for which the cause is unknown.

A child psychiatrist usually doesn't study childhood-onset schizophrenia, but we
have been recruiting a rare form of the disorder because it was a unique way to
study brain development during an age when there are major organizational
changes. I think it should be obvious to people here that since the typical age of
onset of schizophrenia is the early twenties, that obviously you won't know what's
happening in the brain in adolescence unless you select this rare early sort. In
addition, early-onset populations are traditionally a very good way to look for
genes, but I won't go into that in very much detail.

The point of this is just to say that, as seen for adult disorder, our children who had
the mean age of onset of schizophrenia of age ten had a smaller total brain, even
at the time we first met them, had larger brain ventricles, and these represent
findings in adult patients. But where we depart from the literature on the adult
patients, as you'll see, is that until fairly recently schizophrenia in adults was hypothesized to be an early fixed lesion without progression. One of the things that was clear to us from our child schizophrenia population is that they had all sorts of problems before they actually became psychotic at age 10, and that a very . . . this Venn diagram indicates that a large number of them had overlapping problems of social and language and motor functioning, such that a fair number of them were evaluated for autism early on. And while autism and schizophrenia, I think, are well-validated, distinct disorders, for whatever reason early in their development, there were many similarities. And in fact about 20 percent might be considered to have still some pervasive developmental disorder together with autism. But the real point of showing you this is that it's a sign to us that there was a worse hit, so to speak, to early brain development than the typical adult-onset patient had even earlier, before they became psychotic.

And this is my diagram now to show our data set, now considerably increased, for the childhood-onset population, and not shown here is a comparable set for the brain development of their healthy siblings, who are also under study.

Now there have been all sorts of models, hypothetical progression of brains, to do with degenerative disorder and developmental disorder, and this slide is just simply to make it easy for me to say that there have been hypotheses about tissue loss representing an access over the mean, perhaps representing neurodevelopmental disorders. Whereas if it's excessive, it happens later, it's a degenerative disorder. Much of what's being learned, I think, entirely blurs this distinction, and what I want to add is that there's a whole other line remaining to be written here that I think are diagnostically distinct.

So again using the Montreal program initially, what we did was we looked for change, and this is a very crude measure of just total gray matter—frontal gray, temporal gray—in our prospective studies of the schizophrenic children, compared to the yellow being community controls, the normative group that I talked about first. And the blue is a very important contrast group because they're other children, who we're also following, who have transient psychotic symptoms but mostly behavioral dyscontrol and, most importantly, were on the identical medications throughout these years, so that they represent a treatment control. And the point here is that across several areas, but curiously particularly in the back of the brain with the parietal gray, what we were seeing was a very large percent change across adolescence, to a remarkable degree that clearly is of a level that you might even see in Alzheimer's disease, so we were very careful to make sure we were actually doing this right. And, in fact, there have been a few slides . . . this is something called an effect-size comparison, where the red arrow is the size of what we were seeing in adolescence for this group . . . but there have been a handful of studies with adult patients, and the point is that it's very little when you do see progression, relatively speaking, in older patients with schizophrenia.
But here's where a collaboration with UCLA came in, because a method that lets you align brains so that they look more like real brains and pictures of brains allows you to actually visualize changes in brains because you can line up the invariant sulci and then you can actually subtract brains from one group to another. And what we found is that when we looked at the earliest development compared to their yoked control group, you see the red, which represents tissue loss relative to the controls, is mostly in the back of the brain, whereas later at the end of adolescence, you see something actually quite different. And this was the beginning of my career as a moviemaker at the NIH. In childhood I, like probably many of you, were fascinated that you could grow plants in ultraviolet and infrared light, and as a result make movies of plant growth. And even though it takes us 15 years for us, and only two or three weeks for the narcissus, we have actually been able to—15 years later—reap the benefit of these various projects that I've been mentioning and actually study what happens when a child develops schizophrenia. And if anyone ever doubted that this is a brain disease, this is not what's happening in the medication-matched controls, not being shown, and interestingly the area that by the end of adolescence in fact looks the healthiest still is that area, the dorsolateral prefrontal cortex, is about the one that in adults people make the most fuss about, but that isn't where it started. And this was really an arresting sight. And on the medial side, it's not quite the same, it's like a curtain dropping. So remember these two patterns: the back-to-front wave of loss and the curtain dropping from the medial view.

Now, of course, a question that would, I think, occur to most people is, How does this compare to the normal rate? Well, no one ever made a movie of normal development, and that poses certain challenges because with the schizophrenia map that you just saw, we could at each stage, every two years, subtract our patients from their controls, and then in Hollywood cartoon style make a movie of the changes. Well, the controls are the controls; you can't subtract them to anything, from anybody, but you can relate them to an absolute level, say at the age of 4. And so then you can go on to create a movie, and don't be confused here by the fact that as they get old, between 5 and 20, it's blue. It's that the normal movie looks so much like the schizophrenia movie except at a much lower level that I reversed the colors in order to keep them straight. So for the normal movie, remember that blue is loss over time, whereas for the schizophrenia movie . . . so on the left top is the healthy control movie across this age with a back-to-front wave of loss, and that's the schizophrenia one that's just gone. Now notice the bottom left is the medial normal movie with something of a curtain coming down, and then on the bottom right, which I think it will do by itself . . . well, you get the idea. So of course this is now a second interesting fact, and with my captive population of molecular geneticists in this audience, something a child psychiatrist doesn't always get to have, I think this is very tempting to talk about how the game might be up on a normal process, possibly related to synaptic development, that is out of control in the schizophrenic sample.
Gene Expression in Child-Onset Schizophrenia

Another question, of course, was how is this related to genes. Not going into any detail, but, of course, it's tempting to look at the various candidate genes which have recently been so surprisingly and increasingly well-replicated in schizophrenia. And what we're finding is that in fact a number of them, a surprising number of them, give a diagnostic signal within our not terribly large child sample, since we only have eighty-odd probands, we have all their family members in our genetic studies of schizophrenia. But what was particularly interesting to us was that the gene GAD1, which codes for GAD67 expression, that a series of postmortem studies have shown decreased expression of GAD67 in the brain, and, in fact, this is one of the genes found from brain-expression studies, not from positional cloning. And, in fact, the overtransmitted allele is, in fact, associated with steeper decline, particularly in the frontal lobe. But, and here comes my next message here, that before we leap to thinking that we know what these things mean, what we've been struck by was in fact . . . here I'm showing you this incredible tissue loss, and this is where the clinical end of the study comes in . . . the pink part, the right-most part of this age curve, which is just giving full-scale IQ across the study, and starting with the pink color is where they took part in the studies, that there's no loss of IQ for our population in spite of these horrendous change. And much more interesting is that by our four most typical measures of symptoms in schizophrenia—positive symptoms, negative symptoms, overall functioning, total symptoms—for our subjects, in fact, that the gray-matter loss was associated with clinical improvement.

Now this is one of the reasons why I think the clinical part of these brain changes is so important, because when you look at postmortem brain changes in any disorder, you never know what may be restitutive or plastic responses to abnormalities. And it's not too great a stretch of the imagination to think that since in normal adolescents presumably—at least we all like to think that it was the less-used or possibly malfunctioning synapses that get eliminated—it may well be that this is simply an increased reaction to the numerous malfunctioning processes likely to be going on in schizophrenia. This is, of course, wild speculation, far beyond the level of the data that I'm collecting. But I think it's very important, since when we first published this, we were contacted by a number of pharmaceutical firms wanting to know should this be a target of treatment, and the answer is, I don't think you can tell.

We were relieved to find that this loss—and the red is again the loss, the rate going down compared to healthy control group across adolescence—is that this loss plateaus and that, in fact, by the time these patients are 30 or so, they are unmistakable in terms of brain imaging from your typical adult-onset patient, and they're unmistakable in many other ways, including all the family markers of heritability and many biological and psychological markers.
Conclusions

So to summarize this, there's this progressive back-to-front wave of loss. It may be an exaggeration of the normal developmental pattern, and I would love speculations from geneticist colleagues in the audience. It is associated with candidate gene risk but may have a restitutive function, and that we are very interested in . . . and this is a trailer for a movie, not one I'm going to show because it isn't ready yet . . . but we're very carefully following the functioning of the healthy siblings, whose scans we also have every two years in order to see whether or not any differential decline—and I think there will be some—is related to either genetic risk or their functioning, and if it is related to their functioning does it go in a good or bad direction?

[A portion of the transcript including unpublished results has been removed at the request of the speaker.]

But the main point is that, I think, we've come a long way to bridging what child psychiatrists worry about and live with every day and model systems in our head and relating to preclinical work. I think it validates to some extent the DSM people. It's become very fashionable to say very critical things about ICD and DSM, but in a crude way, I think, it shows considerable validation for these unknown and as yet largely unexplained disorders. And I think clinical studies are going to be very important to interpret brain abnormalities, many of which are probably plastic studies, and that these are also proving very useful and phenotypes in our own genetic studies.

And thank you very much for your attention.
Introduction by Richard Mayeux

Richard Mayeux: Okay, we'll move on to the second talk for this afternoon. Huda Zoghbi is a professor of pediatrics, neurology and neuroscience, and molecular and human genetics at Baylor College of Medicine. She's a Howard Hughes investigator. She received her medical degree from Meharry Medical College in Nashville, Tennessee. Her postgraduate training included residencies in pediatrics and pediatric neurology, and a fellowship in molecular and human genetics at Baylor College of Medicine.

Her interests include using tools of modern genetics to understand proper development of the brain, as well as what goes wrong in specific neurodegenerative disorders. She has published seminal work regarding the molecular basis of late-onset neurodegenerative disorders known as spinocerebellar ataxia. Her work in neurodevelopment has led to the discovery of a gene called math1 which governs the development of several components of the proprioceptive pathway, as well as hair cells in the inner ear. Understanding this gene's function has opened up new possibilities for treating hearing loss and for disorders of balance, both of which are widespread health problems.

She's a member of several professional organizations and serves on the editorial boards of many prominent journals. She's a member of the Institute of Medicine and a fellow of the American Association for the Advancement of Science. Her topic today will be "Rett Syndrome and MeCP2: Steady Development."

Features of Rett Syndrome

Huda Zoghbi: Thank you, Richard. It's a pleasure to be here, it's truly an honor and a pleasure. And this morning as I listened to Gerry, Tom, and Eric and Richard each saying how much they enjoy being here, I have to add that even if you're not at Columbia, you enjoy interacting with Columbia scientists and the wonderful neuroscience activities that go on here. And I think I've come enough time[s] back to Columbia which is a measure how much I find it stimulating to visit, as evidenced by a recent e-mail I got from an administrator in the Columbia payroll telling me I'm already in the system and I don't even have to register. This tells you how much I love being part of this, and you can consider me an extended family member and a fan of Columbia.
Anyway, today what I'd like to tell you is about this specific disorder and our approach to study this disease, to really gain insight into both normal postnatal development and the molecular pathways that take place, and also to perhaps learn about broader class of childhood disorders that typically affect children after birth. And that's really why I got interested in this disease. We all have seen children learn different developmental activities, they learn how to sit, they become coordinated, how to walk, speak, often multiple languages, and all that happens within the first two or three years of life. And most of human disorders that are developmental affect the brain in utero, but there are a few rare disorders that actually affect development after the child is born, and that normal developmental program is interrupted. And Rett Syndrome fits into that category, and it was that that really drew me to study it.

What I'd like to do just to give you an idea what the talk is about is tell you how genetics does help us understand a human disease, and not only at the point of identifying the molecular basis of that disease but also at the point of using animal models, even across different species, to finally try to get at the molecular mechanism of the disorder. And I will just go through a few examples throughout the talk just to illustrate these two points.

So girls or patients with Rett Syndrome start life of very typical, they learn how to walk, how to speak, sing songs, and do everything that normal children do. But somewhere between 6 to 18 months of age the Rett girls stop learning new skills and lose already learned skills. So if they were able to speak they will lose the language, and you will see from the clinical feature I'll show you they actually develop many abnormal motor functions. The head growth will slow down, the language skills will disappear, and typically girls when they are born they know how to use their hands very effectively, they can pick objects, they can feed themselves, all of that will be lost and instead will be replaced by these unusual stereotypic hand movements you will see in a film. They will become socially withdrawn, develop autistic features, their balance is abnormal, and they will develop seizures. They also will have tremors, display some anxiety, their breathing becomes irregular off and on, particularly if they're stressed, and they have cold hands and feet, that's why we call autonomic dysfunction.

The next two films will show you some girls with this classic syndrome manifesting many of these features, and this first one was actually the patient I saw as a child neurologist that got me interested in the disease, and you'll see her wringing her hands. This girl, she used to sing nursery rhymes, very interactive, very playful, and at this age of five really all she will do is wring her hands, doesn’t communicate at all, her balance isn’t quite good.

The second girl, you'll sometimes hear the breathing that's irregular. She also has hand movements, rocking activity, and you'll see how abnormal her gait is; she
hesitates, she can't initiate walking, very stiff, very unbalanced, and tend to fall backward because she has major problem controlling her balance [inaudible].

**Genetic Basis of Rett Syndrome**

So this disorder is a sporadic disorder, which means in every family there may be one girl affected, all the other siblings are healthy. And most likely the defect, if genetic, it must have happened in either the egg or the sperm of the parent, just one mutated egg or sperm, because all the other siblings will be healthy. So because of that, because the disease is mostly sporadic, it made getting at the genetic basis very challenging, and in fact took us about 15 years actually to get at the gene. The reason being there are a very, very, small number of cases that were familial, in fact just a handful of cases. It's only these four families that we could use reliably to get at the genetic basis of the disease.

So while that was challenging I would like to go over the process we went through just for you to see that even in having such rare families, one still can get at the genetic basis of disease and eventually understand hopefully the mechanism.

You'll notice in these families that in each family there are at least two girls that are affected. And if you look at the two families on the top, you will notice that it's the mother that's the common parent in each of these pedigrees. These are two half sisters that have different fathers. By just looking at the structure of these families one can deduce that this is probably not a recessive disease, because the likelihood of a mutant allele coming from two different fathers is very small, so that puts it in the category it has to be a dominant disease. The fact that the two mothers, you look at the top, are totally asymptomatic would pretty much exclude an autosomal-dominant inheritance, so that allowed us to narrow the hypothesis further that maybe this is an X-linked disorder; it's on the X chromosome, and these two daughters are affected, and we are going to have to provide an explanation why these mothers are asymptomatic, and I'll get into that in a minute.

So we had these two families for ten years, and really it was these two families that we had to work with for that period of time, until in 1997 an additional two families that were described, shown below, by two other groups where you'll see in one family there is an aunt and a niece who have classic Rett Syndrome, but what you'll notice that the mother of the child in the third generation, she has a little circle and a small dot in the circle, which tells you she doesn't have Rett Syndrome but she's got to be an obligate carrier. Her sister is affected, her daughter is affected, so she must have some sort of a defective gene, but she's not symptomatic. In fact, this girl only had learning disability and she graduated from high school. And in the larger family with three sisters, the mother is perfectly normal. What you also notice there are few lines drawn through some squares. These are males that died in their infancy, somewhere between three months and one year of age. Those males were severely affected with some disorder. We didn't know at the time what this was, all we knew they didn't do anything. These
kids lacked [inaudible], could not learn anything, and required respirators often to support them. So these are the families that we had to deal with, and we hypothesized even with the first two families on the top that this must be an X-linked disorder, and perhaps the females are protected because of a process we call X chromosome inactivation.

And to tell you a little bit about this process, as you know males have one X chromosome and a Y chromosome whereas females have two X chromosomes. Things have to be fair, you know, females can't have twice as many genes as males, therefore one of the female X chromosomes have always to be inactivated. So in every cell only one of the X chromosomes from a female expresses its genes, so the female cells in essence are mosaic, half of them express the X from the mother, and half from the father, whereas in a male where there's a single X chromosome it's always that, one X, all its genes are expressed in every cell. So with that knowledge we then hypothesized that probably in these families what might be happening is that the mother who might carry the mutated Rett gene, shown here as a little red line on the pink X chromosome, she will pass that to her two daughters who will have random X inactivation, half of their cells will have that X as active, the other half the other X, and they will be affected. But in her if she happens to be lucky where only the normal X chromosome is active in the majority of her cells, and that sometimes can happen by chance, she would be healthy and will be spared of that disease. So it was with this we were able to narrow the location of the gene from 23 chromosomes to the X chromosome, and then within the X chromosome now we can look at all these girls who are affected and see which portion of the X is shared by these two half sisters and the other half sisters and which are not to narrow it further on the X chromosome. And that allowed us to narrow it to a small portion of the X where there were about two hundred genes, and from there on it became a brute-force effort, just going through these genes and identifying the right one, and Ruthie Amir, as postdoc in the lab, identified the Rett Syndrome gene to be a code that encodes in Methyl-CpG binding protein, which we call MeCP2 for short.

So I think the first point then of this introduction is that really a handful of families even for a disease that's sporadic would allow one to identify the genetic basis eventually, and those families can be very precious and valuable and should be pursued.

**Mutations in MeCP2**

So what does this protein do? As its name implies, it binds methylated cytosine. Cytosine is one of the DNA bases, and when cytosine is followed by another base, guanine, and if that cytosine is methylated that's when this protein binds. And what I'd like to do is tell you a little bit what's the significance of these methylated cytosine in the genome, and what happens when they're bound by this protein.
Let's look at this illustration first. The little balls on a stick, the pink balls on a stick, these are methylated cytosines. If those cytosines are not bound by this protein, you will have a state of chromatin where genes can be expressed, because transcription factors can come and bind, and there's enzymes or proteins with activities that allows them to acetylate histones, so histones will acetylate, these are the little green balls shown on the histones, and the chromatin is open. Therefore the enzyme polymerase 2 can now proceed through that open chromatin and translate the gene, and an RNA will be made. However, if this protein is bound to these methylated cytosines, this protein can recruit additional proteins, SIN3A, I would like you to remember this protein since I'll come back to it towards the end of the talk, and two additional proteins called histone deacetylases. These are enzymes that can remove now acetyl groups from histones, and when the acetyl group is removed chromatin will be tight, and now the polymerase cannot get through and the gene will be silenced or repressed. So in essence what this protein does, it silences the expression of genes and therefore it's a sort of a repressor. So with this knowledge then we have the candidate gene that's mutated in these familial cases. One will go on and test it in additional families, and indeed it was found to be the cause of the disease in the majority of the sporadic cases.

But the surprise came when we found out that there are a lot more phenotypes that are caused by mutations in this gene than just classic Rett Syndrome. We found that patients with mental retardation and seizures, no other features in females have mutations in this gene. We also found cases with autism that have mutation in this gene. These girls don't look anything like Rett Syndrome, even girls with mild mental retardation and even females who are normal. Remember these mothers in those four families that I've shown you, all of these mothers ended up having these mutations but they were healthy.

So what I'd like to do, and I'll show you two videos of two females we've seen in Houston that have mutations in this gene that looked quite different. This is a girl who was diagnosed with autism. She can walk and she . . . you'll notice though she doesn't have much eye contact, she can follow commands. She will point to her body parts, so she knows her body parts, she can communicate with language, but she has minimal social contact. She did fit the diagnostic criteria for autism, and she had a mutation in this gene. And this other girl that I'm going to show you even higher functioning, she actually has some language, you'll hear her speaking, but she has mild developmental delay and has been diagnosed as such because of poor school performance. So she's playful, she's communicative, but she has learning problems. So I'm sure you can appreciate that these last two girls look very different from the first two girls who had really severe and the classic Rett Syndrome. And what we have discovered is that it's really the patterns of X chromosome inactivation that causes this broad spectrum of phenotypes in females.

If you have the typical random pattern, as we see in classic cases of Rett Syndrome, where half of the cells have the mutant allele that's enough to give that
classic Rett Syndrome. But if the patients are fortunate where they happen to have the majority, maybe 85 percent, 90 percent of their cells expressing the X chromosome with the healthy allele they'll have mild mental retardation or autism, and that's what we found in these last two girls. And those mothers or other carriers are totally asymptomatic, we typically find almost all of their cells expressing the X chromosome that's healthy. So this is really important because it's telling us now there are many females out in the population who might actually be carrying mutation in this gene. They may look normal or mildly retarded, but their kids are at risk of having the full-blown syndrome.

The other thing we learned from having the gene is as physicians started looking into families where they have males with unusual phenotypes that there are actually phenotypes in the males that are also different. The males remember have a single X chromosome. What we found that mutations that inactivate the gene completely, such as the one that's shown below the schematic of the protein, that truncates the protein, gives very sick infants that die in the first three months of year or first year of life. And this was the case of all these males that I've shown you in the first set of four pedigrees, those males had severe mutations, just like their sisters who had classic Rett, but because they have a single X chromosome they couldn't make it and they died.

However what we've also found that there are males that have milder mutations, mutations that preserve most of the protein, such as the third line on this diagram where you'll see just the very terminal portion of the protein has been truncated. These males have presented with mental retardation and seizures, balance problems, so they are severely retarded, but they're alive in their forties, and many of them have a normal life span. More so we found that even subtler mutations, mutations that actually just change a single amino acid and that change is not such a dramatic change, what we call a conservative amino acid change, these males have bipolar disease or may have juvenile-onset schizophrenia as well as some mental retardation. So in the males where we have a single X chromosome we don't have the protective effect of another X chromosome; we're learning about the type of the mutations and the different phenotypes we can see, again broadening the spectrum of phenotypes with this disorder.

**MeCP2 Expression Patterns**

So now what I'd like to move on is some of the studies we have done since discovering this gene, and one of the questions we were quite interested in the gene mutation is there since birth, Why does it take a year, sometimes two years, for the phenotype to manifest in these girls? And one of the ideas we had is perhaps maybe where this protein is expressed and when might it be important. So we began looking at the expression pattern of this protein in the nervous system throughout development. And this is a schematic of a human brain at 10-weeks gestation. At 10-weeks gestation there are very few neurons that are mature, typically at that stage only some of the brain stem neurons and the spinal cord
neurons, right here, typically have matured, and very few at the very superficial part of the brain. And only these neurons that have matured have expressed this protein at that time. By 19-weeks gestation there is a layer of neurons that are called the pioneer neurons, they’re the first cortical neurons to mature. They appear and those now are expressing this protein, and we’re finding more cells positive for this protein in the brain stem and spinal cord. And by 26-weeks gestation now the deepest layer, and this is the second layer in the cortex to have mature neurons, will become positive for this protein, and a lot more neurons in the brain stem and the subcortical areas. And right about birth many of the neurons in the different layers of the cortex have expressed this protein, but not all neurons. There are a lot more neurons that you’re seeing in this schematic, and those are still not positive. So by birth, while we’ve seen many neurons positive, not all of them were positive. And this is just an example where it shows you these pioneer neurons positive at 26-weeks gestation as seen by this immunohistochemical labeling, and this one the deep layer at 32-weeks gestation.

What we were surprised to discover is that the number of cells that become positive for this protein continue to increase postnatally up to 10 years of age. And then once we reach 10 years of age finally almost all cortical neurons became positive. So this was quite important because it told us two things. It told us that first neuronal maturation is really dynamic well beyond the typical 3 or 4 or 5 years of age we used to think at least about, because clearly something is changing in the brain up to 10 years of age. And second it told us that this protein, somehow its levels are modulated in these neurons through this time period, which might suggest that this level could be modulated by the experiences and the development activities these kids are going through, the experiences they’re learning, and exactly how the synaptic activity might be modulating the level of this protein is the question we’re interested in now and pursuing.

**A Mouse Model for Rett Syndrome**

To understand exactly what happens in this disease obviously you’d like to look before the disease has set in and you’d like to have an animal model, because it’s very hard to understand pathogenesis of disease when you’re seeing a child who’s had the disease for ten years and you can at most image them or look at them clinically. So for that we generated a mouse model, and the nice thing about generating mice is that you can choose what kind of mutation you’d like to have. And if you recall, the issue of X chromosome inactivation makes studying this disease in a female mouse very challenging. So the ideal thing would be to study it in a male mouse where you have a single X but you also want that mouse to live, so we picked a mutation that we know causes classic Rett Syndrome in females but at the same time preserves enough of the protein so that hopefully the males will be viable.

And we did indeed obtain mice that reproduce all the features of Rett Syndrome. For one thing these mice, just like the humans, start life of being normal. Actually
they're normal for the first five weeks of life, you can't tell there's anything wrong with them. And then, just like the human, they'll develop tremors, they'll become spastic, they have balance problems, they become hyperactive, but that's all progressive, it happens over several weeks. Tremors start at eight weeks, hyperactivity by fifteen weeks, and so on and so forth, they develop seizures, and they have social behavior abnormalities. And we were quite surprised to find this very unusual, this hang-wringing activity that we see in the human patient, which actually is unique to Rett Syndrome. There is no other neurologic disease that I know of that has this very typical activity. We found that reproducible in the mice. I'm just going to show you a video of such mice just to see what they look like. First you'll see a normal mouse, and by the tail you'll see it keeps its paws open, whereas the Rett mouse constantly having this paw wrenching activity. You'll notice also that it's tremulous, which is something we typically see. And we actually can genotype these mice by just picking them up and seeing who does the paw wrenching and who doesn't. It's that easy to tell that apart.

So in addition to that we can look now, can we reproduce some of the more really interesting features if one is to think about studying disorders such as autism or the social behavioral problems? Can one find social behavior abnormalities in these mice? And we've done many tests. I've just selected just a few examples just to show you a couple of things we find about these mice. We tried to evaluate if they have problems with social interaction, and the way to study that in the mouse you'd like to see how one mouse might interact with another mouse. To do that Paolo Moretti designed these cages where you can have a small compartment with a perforated Plexiglas wall where you can put only healthy wild-type mouse, and now it has a larger compartment in which the activity of the mouse can be monitored by human beings to see where would the new mouse that's introduced into the cage spend its time. Would it prefer to be next to the mouse that's in the confined space, or would it be throughout the cage? And you can do that and quantify the amount of time spent in each part of the cage. Of course to do such an analysis first we have to do the tests when our mice are still normally active, before they develop the hyperactivity, and you want to make sure when there is no other mouse that activity is normal.

So if you look here you'll see that the activity throughout the whole space of the cage is pretty much similar between the wild type and the mutant mice; they're equally active and there's really no difference to the pattern of their activity. But if you look once a mouse has been introduced into that confined space, and that's bin number one, so that's where the mouse would be interacting if it's present there, you'll see that really there is some increase in activity for both compared to the rest of the cage, but the wild type mice spent significantly more—have more activity—near the newly introduced mouse rather than the mutant mice. So this is one of the examples we've done to really try to find out is there a social behavior difference between the wild type and the mutant mice.
Another thing we've noticed about these mice that while we notice they have tremors and they have all their neurologic problems, anytime we try to handle them they appear extremely anxious, they're very tremulous, very nervous. So we wanted to evaluate do they really have anxiety-like behavior, and we've done a variety of tests all of which converge to show that they do have anxiety-like behavior.

But I'd just like to show you one test just to show you the kind of tests that are now being developed to assess mouse behavior. This is called an open field test, and what it is, it is an open field. And as you've heard from Eric this morning mice don't like the limelight, they don't like being in an open field, it makes them very nervous, so if you monitor their activity you'll find they try to run away from that, go to the periphery of the box, try to get away from attention. And both mice will do that, wild type and mutant mice. But as time goes by typically a healthy mouse will start feeling comfortable, realizing there is no threat in that open space, so will start exploring the center of the open cage, whereas the Rett mouse will do much less. And if you were to quantify the data, you'll see that in the first ten-minute interval both are running to the periphery, but in the second and third intervals the wild-type mice spend now more time in the center, whereas the mutant much less so. And of course you confirm that using additional behavioral tests which we have done which told us that these mice do indeed have increased anxiety. And we actually went back and revisited the patients, and that turns out to be quite a prominent phenotype which we had not recognized, everybody focused on the motor abnormalities and language. It turns out that anxiety is indeed a major phenotype.

**Molecular Studies in Rett Mice**

But at least observing such phenotypes allows us now to investigate and ask specific questions: What's the molecular basis of this anxiety? Could drugs that reduce anxiety help this phenotype? Would it be if we treat these animals before they become symptomatic and decrease the anxiety, they might go through the course of their illness, improve their outcome? So such studies are ongoing in the animals in the hope if we find a benefit we can at least treat one symptom that makes the life better for the patients.

We're also trying beta-blockers because these do decrease anxiety. Many people who have stage fright will take beta-blockers before they stand up to give a talk, so that could be helpful. Those drugs also decrease tremors so that might help the animals. So these are examples of things we can do now that we have a good animal model for this disease.

The second question of course one has to ask, okay we have these features, but What are the molecules that are mediating these phenotypes? So to get at the molecules we had a hypothesis, and that is we knew that this protein can bind to methylated cytosines, and as I mentioned to you recruit enzymes that remove acetyl groups from histone, therefore keep chromatin tight and genes repressed,
so the idea was if the protein is mutated now, it will not bind its partners, and perhaps histones will be hyperacetylated and some genes that should be silenced will not be properly silenced. So if that was the idea then the prediction would be that histone acetylation might be increased in this animal model, and you would predict that the highest histone acetylation would be apparent where this protein is most abundant, and will be less apparent in tissues where it is of lower abundance.

So we picked three tissues, we picked the cerebellum because this protein happened to be not as abundant in the cerebellum as it is in the cerebral cortex, but we also picked spleen which is a peripheral tissue where it's highly abundant. And we looked at the levels of total histones, which are the lower panel which shows you by H3 in the wild type and mutant, and those are equal in all three tissues. But when we looked with an antibody that's specifically recognizes acetylate histones, we found that there is indeed hyperacetylation of histones. So it means the hypothesis that this protein is affecting chromatin and histone acetylation is supported by such data. And the question now, Which genes are being misregulated due to that abnormality?

So to do that one can use microarray technology, using again the wild type and the mutant mice, and one will collect tissue from these animals. And what we are doing because of the variety of the phenotypes in these animals we're actually dissecting specific brain regions, regions that are important for movement, such as basal ganglia, or poor cognition, such as the hippocampus, and so on and so forth. And one can make cDNA and label it with a red dye, and the same can be done from a wild-type animal and label it with a green dye, and then put that RNA on a chip that has a variety of genes, just representation of 20,000 or more genes. And the prediction would be, since you put both the red and the green dye, that for genes that are the same level they will show a yellow signal and these would be equal, but those that are higher in the mutant mice will show a red signal, and this way we can begin to identify genes that are altered. And I can report that we have hundreds of genes that are altered, but it takes a lot of data mining and analysis to really get at those that are biologically significant. And that work is in progress.

[A portion of the transcript including unpublished results has been removed at the request of the speaker.]

**Ongoing Work**

And in conclusion I'd just like to end up with this slide which sort of brings everything together. It's really humbling to work on Rett Syndrome because you realize I can't take a reductionist approach, it's a very complex disorder that has many, many features. We can break things down, we are making knock-outs of paminergic neurons, or serotonin-making neurons, so we can study which systems actually contribute to which phenotype. So we have to keep breaking it down in animal models to these simpler or subsets of neurons, but we hope, I think, from studying it we can now link some of these observations to which neurons mediate
the component of movement abnormalities, which ones are responsible for the anxiety, which ones affect the social behavior or autism-like behavior, seizures and cognition. And we're also finding some interesting preliminary data that even altering the environment, manipulating nutrients, can modify the phenotype. So if we really prove that, we can then get at the molecules that are easily affected by nutrients that eventually can affect behavior.

So this is all the work that's ongoing, and in closing I would like to just like to acknowledge the people who've actually done the work. Ruthie discovered the gene, and Mona made the mouse, and the current lab members, I mentioned their work as we went along, Brian is studying the anxiety, Holly made the *Drosophila* model, and Ann has a mouse over-expresser I didn't discuss, Paolo and Juan are mining through the microarray data to get at which ones are biologically important, and Jeff is doing dopaminergic neuron-specific knock-outs. And on the left you'll see all of our collaborators, clinical and neuropathology and Juan Botas, our *Drosophila* collaborator. And finally I'm grateful to HHMI, NIH, and Rett Foundations for funding.

Thank you.

**Questions and Answers**

**Richard Mayeux:** We have time for a couple of questions. There's one in the back. You want to use the microphone or . . .

[Question inaudible]

The question is, How do we go from a mutation in this gene to this explicit hand-wringing behavior? We have no clue, and it's that what's inspiring us to maybe do the knock-out in dopaminergic neurons, because lesioning animals, having striatal lesions in some animal models, have shown that that creates stereotypic behavior. But we have to prove that and we don't have any evidence for that.

**Woman:** Extraordinary story. But in listening to Judith's and then yours, I wonder what other type of brain morphological abnormalities is a factor in girls with Rett, and then I guess you cannot get into the boys because most of them die, right? But is there a pattern that has been described in brain morphology?

**Huda Zoghbi:** No. I'm glad you asked this question. The brain actually looks pretty normal, it's just small, so the head growth accelerates, the brains are 20 percent smaller than wild-type brains. But besides that there's nothing specific. The neurons all appear to be there, there's absolutely no degeneration, so this is not a degenerative disorder. There might be a slight decrease arborization of the dendrites which in the animal models is questionable, it's not solid that one can conclude it's there. So it appears morphologically quite normal.
Introduction by Richard Mayeux

Richard Mayeux: I'd like to start the second half of the program. I'll start off by introducing Professor Sir Michael Rutter, who's professor of developmental psychopathology at the Institute for Psychiatry at Kings College, London. He was a consultant scientist at the Maudsley Hospital from 1966 to 1998, and professor of child psychiatry at the Institute for Psychiatry from '73 to 1998. He set up the Medical Research Council Child Psychiatry Research Unit in 1984 and the Social, Genetic and Developmental Psychiatry Research Centre ten years later, being the honorary director of both until October of 1998. His research has included the genetics of autism, study of both school and environmental influences on children's behavior. He also has a special interest in the interplay between genetics and psychosocial risk factors. He's led a major study into the effect of severe deprivation on Romanian orphans adopted into Britain. He's now entering the third phase in which subjects were followed up to an age of 15. He's well published and is currently the deputy chairman of the Wellcome Trust. Dr. Rutter.

Features of Neurodevelopmental Disorders

Michael Rutter: Well thank you very much indeed. It's a great pleasure to be here. The organizers may feel they've invited me here to give a talk, but I feel I've been invited here to expand my science education, because today's talks have certainly done that.

I'm following on in a way that I hope will show some continuity with what's gone before. The approaches are somewhat different, one in that I'm taking a group of different disorders that seem to have some features in common in order to look at what we may learn from them. I will be asking some questions along the way about continuities and discontinuities between normality and disorder. I will be bringing in a range of areas of science, not just genetics and functional imaging, but also epidemiology and cognitive neuroscience. And I will be emphasizing, and indeed several of the speakers have, that science should not be viewed as a collection of things we know, as a knowledge base, but rather as a process, as a means of solving problems. And accordingly I will be emphasizing the puzzles and challenges that remain as much as the achievements so far.
So let's just look at the concept of neurodevelopmental disorders. They involve delays or deviance in maturationally influenced psychological features, in other words, features that cannot develop unless the necessary neural structure is available. They are disorders, but unlike most psychiatric conditions the course of disorder is not marked by remissions and relapses. The impairment on the whole lessens with age, and that is a characteristic of this group, but the disorder often persists into adulthood, and so it's not just a normal variation. The disorders involve in all cases some degree of specific or general cognitive impairment, and there is a tendency for overlap among these different neurodevelopmental disorders. They have things in common, but they also have some quite striking differences. In all cases the genetic influences are quite strong, but there is evidence that environmental factors are also contributory. And lastly the disorders all show a marked male preponderance.

The four case examples I'm using are specific language impairment, dyslexia, ADHD, and autism-spectrum disorders. That's not all the conditions one could put in this group, but it's the ones I'm going to talk about.

There are things that are reasonably well established. The diagnoses at the core have been shown to be reliable and valid. We know there's a substantial degree of persistence into adult life, and moreover we know that each of these conditions carries important risks for other forms of social and psychological malfunction. And we know that there's a degree of specificity in the associated cognitive deficits. But—I've had to move into a smaller print here—there are a huge number of matters of continuing uncertainty. So that, although the core diagnoses are well established, the boundaries of all four of these conditions remain quite uncertain. Because of that it's unclear whether the disorders should be conceptualized as primarily categorical or dimensional in nature. And that dilemma remains even though there are, for example, imaging abnormalities of a well-documented kind in all four cases. Although there is, through imaging and other modes of investigation, consistent evidence on the existence of neural abnormalities, the evidence is surprisingly inconsistent on their nature, which forces one to think as to what sort of disorder we're dealing with. Fourthly, although there is consistent evidence on the strength of genetic influences, definite susceptibility genes have not yet been identified. And fifthly, there are major uncertainties on the importance and nature of environmental influences. These are, with a few exceptions, multifactorial disorders, so that we know there are multiple genes involved, and the expectation is of multiple environmental influences as well, but we have very limited knowledge on just what those might be.
Research on Cognitive Features

Well just a few examples of findings. I pick on reading disability because there were claims for a while that this wasn't actually as common as it was supposed to be in males. This is a bringing together of four very large epidemiological studies, each of which shows that there is indeed about a 2-to-1 male preponderance. And that is true of other studies as well.

Now let's focus on autism for a moment. There is a lot of research showing not only the presence of cognitive features in autism but also evidence that these are pretty basic to the nature of the disorder. The problem, however, is that we have half a dozen of these. The theory-of-mind deficit, about which I'll say more in a minute, is undoubtedly the best-established, and which is the strongest and most specific of the corollaries, but a lack of central coherence, which I will comment a bit more about in a minute, is also quite strongly associated.

Much less investigated in terms of its implications for autism are splinter skills or idiot-savant skills. The individuals, who, while generally handicapped in their mental functioning, are astonishingly good at some individual skill. These are not entirely confined to individuals with autistic-spectrum disorders but they are much commoner in this group than in any other group of conditions, and if one understood what this meant, it would, I think, provide important clues to the condition as a whole. In addition there are pragmatic language deficits, there are, particularly in early infancy, joint-attention deficits, and, rather less specifically, executive-planning deficits.

So what do we mean by theory of mind? So I go back to the classic experiment published by Baron-Cohen and colleagues in 1985 in which you have two figures, Sally and Ann, and one of them places her object in this basket. She then goes out for a walk, and then while she's out the other child takes the object from the basket and puts it in her own box. And then the question is, When the girl who's been out of the room comes back, where will she look for her object? Will she look where it was when she put it there, or will she look where it in fact now is? In other words, can she think into the mind of somebody else? And I'll come back to findings on that in a minute.

The central coherence is a quite different notion which deals with the fact that all of us tend to look at pictures from an overall gestalt point of view, and we find it quite difficult to break down gestalt to look at something that isn't part of the overall impression. So this is an old test, embedded figures test, in which you see the pram at the top and the question is to find the triangle, or in the bottom a rocking horse to find the house. And the findings on this are that individuals with autism are much better at that than are normal individuals.
Now the association of these has been particularly with autism, and there's now a large literature on this with highly replicated findings. But a follow-up study by Judy Clegg and others, the group that we first studied when they were young children, who comprised boys with a developmental language disorder involving receptive language but who did not in childhood show any autistic features. These are individuals who are now in their mid-thirties and what you see here is the findings from two tests of theory of mind, the awkward-moments tests, and strange stories, and in each case the developmental language disorder group, this column here, are scoring way below their siblings, and way below general population IQ-matched controls, and those are obviously highly significant differences. So we have here a group of individuals who were not autistic, who had a language disorder, and who then in adult life in terms of theory of mind tests are somewhat comparable. Their deficits are less in degree than one ordinarily finds in autism, but very similar in kind.

We'll come back to that group in a moment, but let me turn to a different research strategy which is now looking at what happens to these individuals over time, starting with autism. And this is a follow-up by Pat Howland and others looking at individuals with autism, all of whom had a nonverbal IQ of at least 70, in other words, within the normal range. And this is showing what they were like in adult life. And you can see there is a small group here who are doing very well, they're living independently, they are having some friendships, and are generally okay. But let's look here. We have here a group, not quite half, who despite a normal nonverbal IQ are basically functioning very poorly in all respects. The groups in the middle are sort of intermediate.

So the question then here is, here we have the group of individuals with autism with the best prognosis, as demonstrated by umpteen studies including our own, and yet two-fifths-plus of them are still severely malfunctioning in adult life. And it's not at all clear why that should be so. In passing I would simply say that their level of IQ within the normal range was of no predictive value. IQ was predictive in the broader group in the sense that those with severe mental retardation did uniformly badly and those with mild retardation were sort of intermediate. But no association within this group.

Let's come back now to the young people with developmental language disorders. These are the group now, Judy Clegg's follow-up, in mid-adult life, and a group remember who were selected for showing no autistic features in early childhood, and indeed as far as one could tell socially were okay. But look at those who had a normal range of friendships, less than half, the proportion who'd married or in a regular cohabitation, less than a third. And the schizotypy score was also well up, compared now with the siblings, all of whom had a normal range of friendships, all of whom had married and had a normal low schizotypy score. So we have a dilemma here of a group defined in terms of an absence of social problems in childhood who have become more socially impaired in adult life, or have they? That's to say they are more impaired in adult life but is that because they've got
worse, or is it simply that the social demands have gone up? Well it’s difficult to be sure, but it seems likely that it’s the social demands that have gone up in terms of intense friendships, love relationships and the like, rather than deterioration.

**Genetic-Research Findings**

Okay let's switch to genetics, what does that tell us? As I've mentioned, the findings on all four of these groups is of a strong genetic influence. This is the combined two British twin studies, looking at the percent with an autism-spectrum disorder in terms of the monozygotic co-twin on the left, the dizygotic co-twin, the bluey-green in the middle, and red the general population. So that the increased risk in relation to the general population or the MZ-DZ difference is enormous. The precise heritability that gives rise to depends a little bit on the assumptions you make about the population base rate, but the estimates would be between twenty and one hundred. So it is a huge increase.

But one of the things that came out of the first small-scale twin study that Susan Folstein and I did was that, although there was concordant for autism, there was also an association with what we came to call the broader phenotype, meaning individuals who did not show the major handicap in the disorder of autism, but in terms of their social and communicative functioning, repetitive stereotype behaviors, were very similar. And if you exclude the pairs that are concordant for autism and focus only on the broader phenotype, what you see here is that the broader phenotype similarly has a strong genetic liability. And so back in the seventies then this was the first indication that we needed to think in terms of the genetic liability for autism as going well beyond autism as it was conceptualized at that time. This is looking at a family study. Here of course you can't be sure what's genetic and what's environmental, but the point is you land up with exactly the same pattern. Comparison here between families with a proband with autism and a family with Down syndrome proband. And that whether you're looking at autism or the broader phenotype a big difference between the two.

What about the other conditions? Well this is the twin study undertaken by Dorothy Bishop and colleagues and she approached it in the same sort of frame of mind as we did, focusing initially on serious specific-language disorder as conceptualized at the time. And as you can see there is indeed a big difference between monozygotic and dizygotic pairs indicating a strong heritability for developmental language disorder. But what came as a surprise at that time was that within the monozygotic pairs to a much greater extent than in the dizygotic pairs you were seeing a whole variety of language and cognitive problems of a kind and a degree that fell well short of accepted diagnostic criteria, whereas this was less so in the dizygotic pairs. So the implication once more is that there is a strong genetic influence, but it is applying to a broader group than the conventional diagnosis.

The same applies to the other conditions, but let me jump ahead to just say a word or two on linkage findings. In terms of autism there are partially replicated linkage
findings for areas on three chromosomes. I say partially replicated because of the usual problem about knowing what you require for a replication. It is quite striking that in almost all cases the half-a-dozen international consortia come up with much the same answer. But the problem is it's not in precisely the same spot on the chromosome, and the strength of the linkage varies.

For dyslexia there is rather more in the way of replication, predominantly for loci on 6 and 15 although there are others as well. If we turn to ADHD, the findings are of at least two genes where there either association or linkage and there is replication and there is relevant experimental evidence in support. So it seems very reasonable to suppose that they are playing a role, but the effects are quite small and that it's difficult to know whether to be discouraged by the fact that the only genome-wide scans haven't actually been able to find these genes associated with ADHD, or to be encouraged by the leads that these are genes associated in some way with the dopaminergic system. I mention that because one of the interests in what one finds with susceptibility genes is not finding the genes, which is no big deal—I mean it's proved incredibly difficult—but having found them it doesn't tell you anything very much until you've got some notion of what they do. But with respect to both ADHD and schizophrenia they do link in with evidence on neurotransmitters.

One of the interests in relation to all of these disorders is whether it might be better to focus on so-called endophenotypes rather than the disorder as such, meaning by an endophenotype something which is defined in functional terms in ways that are genetically influenced, and which are related to the disorder but are not the symptoms of the disorder. And of course in ADHD there's quite a substantial body of evidence on cognitive findings and Castellanos and Tannock have put forward notions of three features that they suggest are good candidates for treating as endophenotypes because there is experimental evidence tying them in with the dopaminergic system and with functional-imaging findings.

**Functional-Imaging Studies**

Which brings me on to say a word on functional-imaging studies. This is a rapidly growing literature and it would make no sense to try and summarize all the findings. So instead let me just give a very simplified version of the rationale for using functional imaging in relation to disorder. Functional imaging actually has a range of purposes, so I'm focusing on one particular approach.

So here in relation to disorders such as neurodevelopmental disorders, where the centrality of cognitive abnormalities is so crucial one needs to choose a cognitive process of interest, mentalizing would be a key example, and then select a task to tap that process, which would be one of the various tests of theory of mind. But we need to choose an easy version of the task so that everyone can be expected to be able to do it, and to select a high-functioning sample with the condition being investigated. So the notion is not to show that the individuals can't do the task, it's
actually the opposite way 'round, it's to focus on subgroups who, provided the task is easy enough, can do the task and ask are they solving the cognitive problem in the same way as normal individuals?

So how do you do that? Well you compare cases and controls with respect to brain activity during those tasks and compared with various control tasks, and the question then is whether cases and controls are using the same parts of the brain in tackling the same tasks. There are various other bits that need to be added into that so that an experimental approach in which you can test whether any differences lie in differences in familiarity with the task as distinct from capacity on the task also need to be done. But let me just leap ahead to the findings.

They're interesting but puzzling. So that face recognition is one of the abnormalities that has been well associated with autism, and in normal individuals face recognition is associated with activation of the fusiform face area, whereas in individuals with autism that tends not to be so. Rather there is activation in areas that would in normals be used for processing objects rather than faces. I should say in passing that the part of the brain used for processing faces is ordinarily different from those for other objects. In theory of mind, normal individuals' mentalizing processes associated with activation of both the prefrontal and the temporal areas, whereas in individuals with autism that is less so, there is some of that. But what is more striking is reduced connectivity between the extrastriate regions and the superior temporal sulcus.

Now two things, as it were, emerge from that. The first is that you don't land up with answers which show that this part of the brain is malfunctioning in autism, because it depends which task you're looking at you land up with different answers. And the second is that it appears that it's not so much that there are various areas that are malfunctioning, but that in some way it's the connectivity across areas that is the striking feature. So that Chris Frith has suggested, as have others, that the early sensory-processing areas activate normally, but the later sensory-processing areas are less active. That's to say the dealing with the initial stimuli is okay but what you do in drawing inferences about those stimuli is the problem. And he speculated that this may mean that high-level top-down signaling fails to modulate connectivity, and that this is possibly due to a lack of neuron pruning accompanied by an increase in head size. As Judy Rapoport mentioned, it is striking that head size in autism is often increased and that this appears to develop during the preschool years. Whether it plateaus is more controversial.

Let me turn now to dyslexia and reading, dealing here with a follow-up study by Maggie Snowling and others, which was taking a family high-risk group, i.e., with a loading for dyslexia in the families and a group of controls, and following them up from three to eight years. And what you can see here is that the proportion who are showing reading difficulties in the family risk group is well above the controls; indeed it's a majority. And together with evidence there's not time to present again, the implication here is that we are dealing with something that is distributed much
more widely than the traditional diagnostic concept would lead one to believe. She has suggested—Maggie Snowling that is—that in familial dyslexia that you have a deficit in verbal-association learning, that's what she postulates as the basic handicap, and it's when that's combined with a deficit in oral language that you actually get the functional impairment in reading so that whether or not reading disability occurs depends on the degree to which both these skills are impaired; it is related to the presence or absence of development of compensatory strategies and also the time and quality of teaching, both in preschool and school.

In terms of specific language impairment it's not quite so clear what is happening, and indeed somewhat puzzlingly the findings from the genetic evidence that Dorothy Bishop has dealt with—not just the study I referred to earlier but also other data—suggesting that there may be a different route for the genetic influences and the environmental influences. For a whole variety of reasons that would be quite surprising, but if indeed turned out to be true would be very important.

Conclusions and Challenges

So let me pull things together with just a small set of conclusions. We start with the evidence that neurodevelopmental disorders constitute an extremely important and interesting group of seriously handicapping conditions about which we know a good deal. They have two very important things in common, well they have more than two, but especially two. The first is the marked male preponderance, and the second is the association with specific or general cognitive deficits, as well as the general features of early onset, persistence into adult life, and lack of remissions or relapses. So that the question that comes up with the male preponderance is, Is it simply coincidental that the disorders showing a male preponderance are almost all neurodevelopmental disorders, in marked contrast to those showing a female preponderance which are basically adolescent onset, emotional disorders, eating disorders, and things of this kind? The theories mainly deal with one of these conditions at a time, and come up with different hypotheses for each condition. Well I think the bringing together of disorders tells you that of course there may well be, probably are, specificities, but it would be a very strange coincidence if this group of disorders, the male preponderance, didn't have something in common.

The association with specific or general cognitive deficits, although the details of the deficits are different, there is overlap; the theory of mind I mentioned in relation to specific-language impairment and autism, but there are others I could mention, so that the suggestion that I would want to leave you with is that we have various ways of tackling these problems. Some, we've heard important things during the day, approach it, as it were, from the basic science end, some approach it from the point of view of looking at normal development, some from a strong focus on an individual disorder. All of those are worthwhile, and indeed all of those are going on with autism, but in addition I think going across disorders may be valuable.
But in saying that I've got to come back to the fact there are important differences among this group of disorders. And so we have to deal with the question of the commonalities and the differences and both have to be taken on board, and of course the usual question, are these two sides of the same coin or are we dealing with a different set of factors that explain the heterogeneity?

Now in all cases it seems likely that we are dealing with a systems disorder of one kind. I've emphasized that particularly in relation to autism with respect to the functional-imaging findings, but it actually comes up somewhat similarly in dyslexia, for example. But the question then is, What sort of systems disorder? And with the exception of ADHD, what is striking is that drugs have proved remarkably lacking in benefit. There are some drugs that in some children make some difference, but if you compare it with ADHD or depression or schizophrenia or bipolar disorder one has to say, here we have what all of us think are systems disorder, and yet we find nothing of any great interest in drugs that affect neurotransmitters. Well maybe we're looking at the wrong neurotransmitters. But it is still a puzzle.

And so, as I say, it's only in the case of ADHD is there a substantial response to medication, and that does make it different from the other disorders. Of course with ADHD we have the problem, as we have throughout the whole of internal medicine, that we know a great deal about group responses to treatments, and the same applies to psychological treatments or medication, but we know much less about why there's such individual variation, that even with the most powerful drugs there are individuals who don't respond and individuals who do respond, and we have a very limited leverage on what that is due to. Pharmacogenetics may provide an answer, but it hasn't as yet. And of course there are similar differences in terms of side effects.

Genetics I think is going to be helpful in leading us forward in clues as to what biological studies need to be undertaken, and it will be very important that genetic research goes hand-in-hand with the rest of biological studies, in both basic science and clinical neuroscience, because the genes themselves are of interest only once we know what they do at a molecular level, but then is the further issue as to how it leads on from the effects on proteins to the phenotype, and that's an even bigger set of issues and we're at a point I think when we can see the road down which we want to go, but we're quite a long way from seeing where we're going to land up.

And on that point I will end. Thank you.

Questions and Answers

Eric Kandel: I found that just fascinating. I was particularly taken by three of your remarks. One is that this is an inappropriateness of connection that you have in autism, so it's a misconnection syndrome, and contrast it to the disconnection
syndrome that [inaudible] was speaking about to the imaging studies which suggest that when autistic kids look at faces they see objects, and third the lack of response to drugs. So it seems to me this suggests if the misconnection idea is correct that perhaps the defect is really quite specific, and it isn't early stages of visual processing versus late stage, it's the particular pathway that is concerned with object recognition. So one would predict, for example, that these children would have no difficulty in recognizing objects in space or movement of objects in space, that the pathway concerned with where objects are would be intact, number one. And number two is if it in fact is a misconnection syndrome, I don't know why one would expect modulators of transmission to affect it. This presumably is sort of a Tom Jessell problem; it's how neurons find each other that maybe is an abnormality in cell recognition, cell adhesion molecules, or something like that.

**Michael Rutter:** Well I agree. Those are very perceptive remarks, as always. The reason for emphasizing the neurotransmitters is that virtually everybody has assumed that's where the answer lies. But I agree with you that for the reasons you put forward that actually doesn't seem very likely, and if Tom at the next time there's a symposium, but I hope we don't have to wait 250 years, can find an answer to what is causing these misconnections, that would be great, I will be back.

[Question inaudible.]

Well that's a good question, but it's actually a very difficult question to . . . The question is, Of those four disorders which are the ones that are most affected by psychosocial influences, and affected by detriments in self-esteem and the like?

The problem is that you will probably pick out ADHD, I suppose, but the dilemma then is, Are the psychosocial factors which are associated with malfunctioning socially in middle childhood and adolescence telling you about psychosocial influences as they impinge on the hyperactivity as it was present in early childhood, or rather on the transition to the conduct problems and so on that follows? It pretty certainly is the transition to conduct problems. Whether in addition there is a causal effect on the hyperactivity itself I don't know. But let me throw in two findings which were puzzles to me. I've never understood, I must say, why psychologists spend most of their time being firmly committed to showing that they were right. My interest in research is understanding why I was wrong, and I am keen to say, I've had several examples in my career of being spectacularly wrong, but I think I've learned from each of them. The two findings which were surprising is that in our follow-up of Romanian adoptees—these are kids who suffered profound deprivation in Romanian institutions and who were then adopted into UK families—and I wouldn't have dreamt, I have to say, of looking for autism in this group were it not for the fact that at the time we started the study I had two clinical referrals of children from that background who were said to have autism. To cut a long story short, what we found was that there was a substantial increase in what we called quasi-autistic behavior. We called it quasi-autistic because although at
age 4 the pattern was very similar to ordinary autism, by 6 it was somewhat
different, and by 11, because we’ve done the follow-up at 11 now, was even more
different. So the question then is, Is that telling you something about the cognitive
abnormalities in autism, or is this an entirely different phenomenon? And I don't
know what the answer to that is. I think we have to bear in mind both possibilities.

The other surprise also from the Romanian study was a substantial increase in
ADHD-type problems, although again they were somewhat atypical, so that it was
much more on inattention; they weren't particularly overactive, although a few of
them were, and the question is again, Is that telling you something about ordinary
ADHD, because neither in autism nor in ADHD does one ordinarily find many
children who've been institutionalized, that's not the usual sort of background, or is
it entirely different? And again hopefully the follow-up and the comparisons that
we're now doing may tell us about that.

I think the other thing I've got to throw in, though, is that when we talk about
environmental factors we need to be concerned that what we actually mean are
nongenetic factors, and therefore one's got to take on board the possibility of either
epigentic influences of one kind or another, or development perturbations which,
as it were, are feeding in. So the field is open. I don't think there's a good answer
to your question, I'm afraid.

**Question:** Well, just on a sort of clinical level. I kept waiting for you to include in
your grouping Fragile X Syndrome which seems to me to have so many of the
characteristics and might be helpful in facing some of the questions you ask; we
have the chromosome, we have development, we have boys, and we have
psychological and other interesting sidelines to it. Would you consider it or doesn't
it qualify?

**Michael Rutter:** Time did not allow. The Fragile X I would see actually as
somewhat different, but it has certain features which are very much in common.
That's to say the original claim that something like a third or even half of autistic
individuals showed Fragile X is clearly wrong, and was based on quite inadequate
laboratory methods as they were available at the time. And now that we've got
DNA methods it's clearly less than 5 percent, and maybe a good deal less that 5
percent. But if you say what is the proportion of individuals with Fragile X who
show problems, that overlap with autism then it's much larger. And so again we
have this difficulty of where you draw you line, and that, as Judy Rapoport was
saying, *DSM-IV* and *ICD-10* have been very useful, they're not wholly wrong, but
they don't match well with certainly the boundaries of the concepts that we're now
dealing with, and in some respects even with the divisions amongst them. That's
one of the interesting things, and it is essentially open to research and [inaudible]
will be included on the list.
Brain and Mind
May 13, 2004

Nora D. Volkow, MD
Drug Addiction: The Brain in Disarray

Introduction by Richard Mayeux

Richard Mayeux: For those of you who have persisted through this afternoon, [you] are in for what I think, what's going to be a very exciting talk. Reading through the papers by our next speaker have just sort of illustrated to me the very usefulness of brain imaging in understanding the basic disarray in the brain for disorders of substance abuse and addiction.

Nora Volkow is the director of the National Institute on Drug Abuse, or NIDA. Prior to that she was at the Brookhaven National Laboratory as the director of Nuclear Medicine, and the director of the NIDA Department of Energy and Regional Neuroimaging Center there. She was also a professor in the Department of Psychiatry and the associate dean for the medical school at the State University in Stony Brook.

Her research investigates the mechanisms underlying the reinforcing addictive and toxic properties of drugs of abuse in the human brain. She was the first to use imaging to investigate the neurochemical changes that occur in the human brain during drug addiction. And most recent studies have employed imaging to investigate the effects of stimulant drugs and to examine changes that occur with aging in the dopamine system. She's a recent member of the Institute of Medicine, and was named the Innovator of the Year in 2000. I'm pleased to introduced Nora Volkow, who's going to talk on drug addiction and the brain in disarray.

Neurobiology and Drug Addiction

Nora Volkow: Good afternoon. It's a pleasure to be here, and I want to thank the organizers for having included me in this wonderful symposium.

I'm going to be changing subjects, and this is the brain in disarray, which I call the state of drug addiction, which is actually . . . what intrigued me about it to start investigating drug addiction, which I started 25 years ago, was the notion that in people that are addicted, one of the characteristics that we value the most as humans—that is our ability to choose, to do our behavior and be able to carry it, our ability to control our behavior—is markedly disrupted. The person that's addicted to drugs does not choose to be addicted, and what happens is at the time that they end up coming to see a psychiatric unit they basically are at the point that
they realize that they no longer can control their behavior and they need our help. So it was this paradigm that actually led us to try to understand what happens in the brain of an individual that, even though they consciously want to stop taking a drug, they are no longer able to do that, or it's very difficult for them to do that.

Drug addiction is a disease of the brain for which we have a culprit; we have drugs, we can point a finger at drugs. However, we also realized that the chronic drug exposure, while in some instances leads to addiction, in others it does not, and what has become evident is that there are other variables that are indispensable and necessary for the addiction to occur. A very important one are those biological and of them, genetics have been identified by studies in humans. It has been shown that close to 40 to 50 percent of the disease of addiction is indeed accounted for by genetics. There have been multiple genetic studies, there have been several loci in chromosomes identified and replicated, but there's no gene yet that has been linked directly to drug addiction, except for specific ones that relate to metabolism of drugs such as nicotine and alcohol. But for the most part we do not have the genes that can be linked to the genetic predisposition for addiction.

It is also evident that environment plays an extremely important role, and here I have to be careful because of the previous speaker. And when I speak of environment the variables that come forward as some of the most relevant ones in enabling addiction are those that related to stress, acute as well as chronic stress, and of the stressors that are linked with addiction one of the ones that are stronger are those related to social type[s] of stressors.

So in this paradigm how do you try to understand what are the neurobiological changes that lead to the phenotype of drug addiction. And before I go into that let me make some disclaimers at the beginning rather than at the end. When you are doing studies where you do not know the genetics it becomes harder to target a particular system, number one. Number two, in a process that you're studying drug addiction what happens is you get the individuals when they are already addicted so you do not know the extent to which the changes you are identifying were there before the person became addicted, or whether in fact they are the effects of chronic exposure to drugs with interactions of the other variables. Also the other disclaimer is if you go into the brain, which system do you go into? And we chose to specifically investigate the brain dopamine system, and I made it very clear here, not because we believe that drug addiction is just involving the brain dopamine system, because we wanted to identify what the role of the brain dopamine system was in the process of addiction.
The Dopamine System in Addicts

And why did we chose that in that case, dopamine as a neurotransmitter to go after? It was because there's multiple studies that have been done in laboratory animals showing that basically all of the drugs of abuse that can produce addiction in humans, all of them increase dopamine in the nucleus accumbens. And this ability of drugs to increase dopamine in the nucleus accumbens has been shown to be extremely important in the reinforcing effects of drugs of abuse. So drugs, when they are taken, increase dopamine in the nucleus accumbens and then that is linked to what we have been calling reinforcing.

For many years it was traditionally felt that reinforcement, that is the characteristics that will lead the animal to take it again and again, was equated with pleasure, and many studies for many years equated dopamine with pleasure. For the past ten years of neuroscience research, it has become clear that the dopamine system which is targeted by drugs is not per se relevant, perhaps to hedonic pleasure, as we call it, but actually signals something that is more indispensable for survival and motivation, and that it signals the saliency of a particular stimulant. And pleasure is one of the characteristics that make the stimulus salient. But by the same token an aversive stimulant can also be salient, a novel stimulant can also be salient, an unexpected stimulant can also be salient. And so what it is basically telling us is that drugs of abuse are activating the system that is the way that the brain will process something as salient.

Now one of the aspects that has been said is yes, it is recognized that the ability of drugs to increase dopamine is indispensable for the reinforcing effects, but how does that explain the process of addiction? And it doesn't explain it, because if you were to give a drug—and this has been done in animals and humans that are not addicted to the drug—and compare the response to that of an animal or a person that's addicted, what you find is that the drug is able to increase dopamine in the nucleus accumbens in both the addicted and the nonaddicted subject. In fact, in certain instances the response is even larger in the nonaddicted person. So that ability of the drug to increase dopamine per se is definitely not giving us any clues about what is it that leads the person to lose control and engage in that compulsive pattern of drug administration that characterizes drug addiction. So we decided to do the studies and ask the question, Is the brain dopamine system at all involved in the process of addiction, and if it is involved, how does it help us understand the phenotype?

PET Studies of Addict Brains

So we used positron emission tomography to investigate different molecular elements involved in dopamine neurotransmission. And at the same time we used a strategy—this has been going on for the past twenty years—of investigating a wide variety of addictive disorders. So we started alcoholics, we started cocaine abusers, we started methamphetamine, heroin, marijuana . . . and what we've
been interested by studying and comparing all these addictive disorders is to identify if there are common abnormalities of all those diseases, of all of these addictive diseases. And the reason why is we are not per se interested in understanding the specifics that relate to cocaine addiction versus alcohol addiction, but to understand the common characteristics, the common neurobiological processes, that explain the common phenotype, which is this one that is the same in an alcoholic as in a cocaine abuser, which is that inability to control the intake of the drug, even though they consciously don't want to do it, and the engagement of this compulsive behavior of taking the drug again and again.

So that's what we asked, the question, What elements in the dopamine neurotransmitter system are involved, if at all, in drug addiction? So we investigated, in the same subjects, we investigated three different elements that are involved in dopamine signal neurotransmission. One of them is the dopamine transporters, which we were using as markers of the dopamine terminal. This is an image of dopamine transporters. Forget about them, because while they are affected in certain addictive diseases like methamphetamine, these arrangements are basically secondary to neurotoxicity of these drug and dopamine terminals. But this is not found in alcoholics, is not found in cocaine abusers.

On the other hand, this other element, the dopamine D2 receptors, which is one of the dopamine receptors that has been shown to be involved in the translation of the dopaminergic signal induced by salient events into the required centers of the brain, or the motivational centers of the brain. And the dopamine D2 receptors similarly had been linked with the reinforcing effects of drugs of abuse such that if you block them by giving neuroleptics, you profoundly affect the reinforcing effects of drugs such as... basically almost any drug. And also if you create knock-outs that don't have dopamine D2 receptors you profoundly disrupt the reinforcing effects of various types of drugs in these animals. So dopamine D2 receptors are actually one of the ones that I'll concentrate [on] because they are affected in similar ways by a wide variety of drug addictions.

And finally, in the same subjects we measure brain glucose metabolism because under normal physiological conditions, the main source of energy for the brain is glucose. So then by measuring the different levels of glucose consumption you can actually get an indication of how active the brain is, and it's very sensitive actually to dysfunction, to the arrangements in brain activity. So I'm going to concentrate on dopamine D2 receptors and brain glucose metabolism.

**Dopamine D2 Receptors in Cocaine Abusers**

What did we find in dopamine D2 receptors? This is one of our first studies. It was done in cocaine abusers that were hospitalized to ensure that they were not taking the drug. And they were tested with spiroperidol. All of these are males. And first I'll show you the images and then I'll show you the numbers, the individual numbers. This is the scale to the right, and methylspiroperidol is a ligand that binds both to
dopamine D2 receptors and serotonin receptors. But this binding to these receptors has different pharmacokinetic properties. So by waiting, say, two hours, you can eliminate mostly the signal that is coming from serotonin and your signal reflects dopamine D2 receptors. So these are different levels of the brain going at the planes where you have the basal ganglia from the upper part to the lower part, and in the posterior aspect here is the occipital cortex part of cerebellum where you have no activity because that area is, for practical purposes, devoid of dopamine D2 receptors. So this is a normal subject. You see the high activity into the basal ganglia.

Look at the cocaine abuser one month after last use of cocaine, and you can, just by looking at the images, see a significant reduction in the binding of the radioligand. And this reduction persists long after the patient has stopped taking drugs. And in this case we kept the patients for four months in the hospital to ensure that they in fact were not taking drugs. And we documented effectively cocaine abusers had disruptions, decreases in dopamine D2 receptors, that are long-lasting and certainly persist at least after four months of last use of the drug.

This is an image . . . the next slide basically shows the data but in a quantitative fashion. So you draw a region, you actually measure the concentration of your radio tracer and do appropriate mathematical modeling that then allows you to extract an indication of the receptor numbers in that particular area of the brain, or what we call receptor availability.

And so here it is, the measure of dopamine D2 receptor availability for normal controls in pink and for cocaine abusers in green. What I'm doing is I'm plotting it as a function of age because as we grow older the expression of dopamine D2 receptors is significantly reduced, approximately 4 to 6 percent per decade. So you see that reduction in normal controls, but you also see it in cocaine abusers. And if you look at the data as a whole you can realize that as a group cocaine abusers have significantly lower levels of dopamine D2 receptors. And this was significant. We've replicated that same finding in three different groups of cocaine abusers. Here at Columbia University, Marc Laruelle and his group has also documented the same finding. Others have used SPECT and [inaudible], also documenting reductions in dopamine D2 receptors.

However, I want you to look at the data and realize that this is not a categorical distinction. And it's something that has always intrigued me in terms of what does it mean to have a finding that it is statistically significant, but where there is overlap between your disease population and your normal controls? And, for example, look at this subject here with the normal control who indeed has lower levels of dopamine D2 receptors as those levels similar to those of cocaine abusers. So when I get data like this the way that I interpret it is to say that while as a group cocaine abusers do have lower levels of dopamine D2 receptors, and this in turn may tell me something about propensity for taking drugs and addiction, it certainly is also telling me that it is not sufficient to account for addiction, because if it were
sufficient to account for addiction then how do you happen to explain a cocaine abuser with levels that are normal or a normal control with levels that are lower? And I'll come back to that statement because I think it gives us an insight into what the significance of the differences in the receptors are with respect to addiction.

The other thing about these findings of decreases in dopamine D2 receptors in cocaine abusers that, as I say, is a very replicated finding, it is significant but it is not perfect, it's not categorical. The other aspect about it is it's not specific for cocaine. We and others have documented that dopamine D2 receptors . . . these are controls, this is just an illustrative pair of a control and an alcoholic. Measurement is with C11-raclopride, a ligand that binds to dopamine D2 and dopamine D3 receptors, the scale for dopamine D2 receptor availability. So we and others using different types of radio traces have documented that alcoholics that have a family-type history of alcoholism also have reductions in dopamine D2 receptors. We have also shown heroin-addicted individuals have decreases in dopamine D2 receptors. And we recently documented the same thing for methamphetamine addiction. So across a wide variety of addiction, excluding nicotine—in nicotine we and, I think, the group here at Columbia, that's one of the addictions where this does not seem to be consistent—but otherwise in cocaine, in alcohol, heroin, methamphetamine, this is a consistent finding.

**The Role of Dopamine D2 Receptors**

Now if you have a reduction in dopamine D2 receptors, and as I was saying, how does it help us in any way identify how this could translate either . . . two possibilities, I'm putting it as two possibilities. Number one, if you have low levels of dopamine D2 receptors, as it appears from the cocaine abusers, you could say could this be making you more vulnerable? It's not obviously per se making you addicted, but it could make you more vulnerable for addiction. But let me be again subversive and say the other completely different possibility, which is maybe it is not that the low dopamine D2 receptors are making more people vulnerable, maybe what's going on is having high levels of dopamine D2 receptors is protecting individuals to take drugs. And let me then go through where my thinking is going here.

So if you have low levels of dopamine D2 receptors or high dopamine D2 receptors, what simple explanation can you make? Well, when you have . . . this is a dopamine synapse . . . I mean the dopamine cells are not there for us to take drugs. The dopamine cells are to signal saliency. So the probability when the dopamine cell fires in response to a stimulant, the magnitude of how intense the stimulant is will determine the magnitude of how much dopamine is released. And then there is a probability of dopamine interacting with receptors. And that probability is not infinite, is basically quite fixed. And it is fixed because dopamine is liberated from the synapse but it's immediately brought back, is recycled back, into the terminal by these dopamine transporters. And mathematicians had modeled actually how long those dopamine stay in the synapse, and it has been
estimated that any given molecule of dopamine does not stay in the synapse for longer than 50 milliseconds. So the time at which dopamine has to interact with the receptors is limited. So then the question that follows is for a given . . . the same type of increase in dopamine what basically will determine the probability of an interaction for the same amount is going to be determined basically by the number of receptors.

So if—as in a question like this one—you have even a dopamine signal that is very mild, you have a lot of receptors, the probability of an interaction is quite large. On the other hand, if you have the levels of receptors are decreased the probability of an interaction is decreased. So this is the way that we simply try to understand what the significance will be of having a decrease in receptors. The way that I would view it is it would make you less sensitive to the signaling of stimulants that are salient, because the probability of an interaction would be lower. And I'll come back to this and how then this can make someone more vulnerable, if it is the case of vulnerability, or alternatively if it's protective.

Now one of the aspects that I started with is that in the neurobiological changes that were documented in addicted people, we really do not know if these were there before they starting the drug. We don't know if the receptors were already there or whether chronic drug exposure leads to these decreases in receptors. And this is not a trivial question because it pertains, to me, to one of the most challenging questions that we face in the field, which is to understand why people, when exposed to the same drug, to the same environmental conditions, some become addicted and others do not. And it pertains very much to issues of genetics as well as biological developmental exposures that you may have encountered while growing up.

So how do you address this particular question? Well of course it would be ideal to take people and measure dopamine D2 receptors before they become addicted, and then to actually follow them and then wait until they become addicted and test them. But you may realize that that basically is a very impractical proposition to start with because drug addiction starts in adolescence and it is basically very difficult to use this type of nuclear-medicine technology to image adolescents. And moreover the cost of such an experiment would be prohibitively expensive. So how do you tackle it? And we decided to tackle it in a different way. This is . . . basically what you're seeing here is a scattergram for another one of the studies. I've told you we've replicated the studies comparing dopamine D2 receptors in normal controls and cocaine abusers in three different groups of subjects. And we do that with imaging because with imaging technology you can not gather large numbers, so you need to replicate a particular finding to ensure that it was not just due to a sampling of subjects. So this is one of those studies. These patients were also inpatients. They were studied three weeks after the last use of cocaine, and you see in this case it's a C11-raclopride study. You see exactly the same finding: cocaine abusers as a group have lower levels of dopamine D2 receptors. But I want you to basically again see the overlap. Look at these normal controls whose
values are really undistinguishable from data from [a] drug-addicted person. So I want you now, having mentioned this fact that this is very similar to this of the cocaine abusers, I want you to forget completely the cocaine abusers and look at the normal controls. And what always has fascinated me about the normal controls is the wide variability in levels of dopamine D2 receptors that we observe in healthy normal controls that are not addicted to substances. And even though some of that variability is accounted by age, you can also see that it is not the only variable explaining the variability. Look at these subjects, their years of age, and this one here has 50 percent higher levels of dopamine D2 receptors.

Receptor Levels and Drug Response in Nonaddicts

So we decided to take advantage of this variability in the expression of dopamine D2 receptors to ask the question, If indeed we're postulating that levels of dopamine D2 receptors, either by being very low makes you more vulnerable, or alternatively by having high levels may protect you, then what does it mean to be a person that has low levels of dopamine D2 receptors vis-à-vis the way that you respond to a drug of abuse? And this is a nontrivial question because, if we are indeed seeing and implicating these receptors in drug addiction, it follows that it should have an effect on the way the levels of expression of these receptors it follows, we hypothesize, should have an effect on the way that people respond to drugs of abuse. So that's the way that we decided to address the issue of the extent to which differences in dopamine D2 receptors before a person becomes addicted in any way affects their responses to drugs.

We did a very simple experiment. We again replicated it twice, but the first one, which is the one that I'm going to be showing you . . . we took 23 healthy controls, we measured levels of dopamine D2 receptors, then we brought them back to the laboratory and we injected them with intravenous methylphenidate. Intravenous methylphenidate—which is basically methylphenidate, is Ritalin, which is the most widely used psychiatric drug in the treatment of attention deficit disorder—is pharmacologically very similar to cocaine, like cocaine blocks dopamine transporters. And when you give methylphenidate intravenously to cocaine abusers, in fact they like the drug very much, and they say it's actually very similar to cocaine, not the same but very similar.

Interestingly when you give intravenous methylphenidate to healthy controls, approximately half of them report the drug as very pleasant, and the other ones report the drug as very unpleasant. So the responses to stimulant drugs in people that are nonaddicted to drugs, whether you call it methylphenidate or amphetamine, is quite varied. And this has been described both in the animal as well as in the human literature. So we decided to ask that question, the question of do the levels of dopamine D2 receptors in the human brain in any way regulate the responses to stimulant drugs? And so we measured 23 healthy controls' dopamine D2 receptors, we injected them with intravenous methylphenidate, and we did a [inaudible] of self reports of drug effects. But our main variable, dependent
variable, was, is the drug experienced as pleasurable or is the drug experienced as unpleasant? That was the first [inaudible]. And what we found was as follows, indeed the availability of dopamine D2 receptors in fact affected the type of responses that subjects were given. And here it is. First look at the images, these are a subject that's reporting the effects of methylphenidate as unpleasant. This is a subject that [is] reporting them as pleasant. The images have been already transformed, so the numbers reflect availability. And you see this subject that reports it as unpleasant in this particular . . . for this comparison of this group, has significantly higher levels of receptors that this one here. And as a group you see this is quite significant. In individuals that are reporting the effects of methylphenidate as unpleasant, and it was very unpleasant, have significantly higher levels of receptors. Those that are reporting it as pleasant are significantly lower. And in life there are always outliers, so I have two outliers to the right that we have to ignore because I don't have time.

But I'm going to come back to this one here and ask the question, so why would that be so if you have such a finding? And the first thing why it would be so is—again I go into the preclinical literature to try to get an explanation—and what you find in the literature in reinforcement and investigation of reinforcing [inaudible], what it has been shown, for example, is if you produce an electrical current into the lateral hypothalamus, the electrical current is perceived as pleasant or reinforcing by the animal—I shouldn't call it pleasant—reinforcing by the animal, and the animal will press a lever in order to deliver the current. But what investigators have also shown is that if the current is too low it is insufficient to generate that response so the animal doesn't bother and does not press the lever. But what's fascinating is if the current is too high the animal stops pressing the lever because the current becomes aversive. So there's an optimal level of stimulation by which the activation appears to be reinforcing and makes the animal press the lever.

So if you use a similar type of analogy it follows that you could say, perhaps the reason why, in individuals with high levels of dopamine D2 receptors, methylphenidate is experienced as unpleasant is because when you inject intravenous methylphenidate, and this has been studied in animals, the magnitude of the increasing dopamine is at least five- to tenfold higher than that of any natural reinforcer. Not only is it significantly higher in magnitude, but that duration of dopamine in the synapse is also significantly, significantly longer. So could it be . . . the way that I would interpret it is in the simple way, the first go-around, is that when you have a lot of receptors when you take a drug of abuse—and drugs-of-abuse characteristics is that increases in dopamine are very, very large, they are supraphysiological—that increase is perceived as aversive because you have so many receptors it's overstimulated. On the other hand, if you have low levels of receptors, those low levels basically are able to blunt the large increases in dopamine, bringing the stimulation to the level, to the threshold, that is perceived as reinforcing.

High Receptor Levels and Averse Drug Response
And this is what we hypothesized. We hypothesized that perhaps the reason why this was occurring was indeed because, in this case, is what the supraphysiological increases in dopamine led to an aversive response. So we decided to test it and we said if our hypothesis is correct then it follows that the reason why these subjects are perceiving this as aversive is that the dose was just too large. What about if we get one-tenth of the dose? So we went to the IRB, we asked permission, and basically the rationale was we want to bring these subjects back, give them one-tenth of the dose, instead of 0.5, 0.05. And we hypothesized that that dose will be perceived as pleasant in these subjects. So we got the IRB approval, which took many months. And so finally we got the approval, we called the subjects and the subjects refused to come back because their response to methylphenidate had been very aversive, it had been aversive to the point that we had to report that to the IRB. So I don't know exactly whether on that level the hypothesis was correct. I think it was, that basically we were dealing with too large of a dose. But it is a characteristic of drugs of abuse, whether the doses that you take it and kids are exposed to, they are inducing supraphysiological responses. In this case low levels basically are not going to be leading these very aversive responses. And so when you, as a kid—because as I told you most drug addiction starts during adolescence—are exposed to a drug and you're exposed and you take it and say this feels good. And then the next time that you are in a party and they give it to you the likelihood that you will take it is much, much higher—and this has been shown by behavioral studies— than if [when] you took that drug the experience was so unpleasant that you refuse to come back. So if the kid gets the drug and basically gets a very aversive response, the likelihood that he will take it again is much lower. So you could then bring this up and say could it be that this is one of the mechanisms by which indeed . . . that having low levels of receptors make you more vulnerable because you will not get an aversive response. So that's what we hypothesized.

Now when you are doing imaging studies like this one you are limited. You are limited because you can not manipulate variables like you can in laboratory animals. So you are limited to the notion of replication. We've replicated exactly the same findings and show indeed that individuals with high levels of receptors have very aversive responses to methylphenidate and this is consistent whether you study them on different days. However, this is an association. How do you assess whether in fact there is a causal link, and what do I mean by a causal link? I mean whether indeed having high levels of receptors is causally linked to having an aversive response. How do you demonstrate that causal link in human studies?

Well the way that you would demonstrate that causal link would [be] you have these subjects in whom you have low levels of dopamine D2 receptors. If they are causally linked to aversive responses, if you increase the levels of receptors then that experience when you gave them the drug should be aversive. The problem is how do you increase dopamine D2 receptors in a human brain in a noninvasive way? And I don't know of any way to do it. However, you can do that in animals.
And so this is an example of how we're blending—which was elegantly illustrated in the talk earlier—human findings to then design animal studies to then try to extract the pertinent information. So we asked the question, We can not do it in humans but if in animals we basically train them and make them addicted to drugs and then in the same animals, then, when they are already self-administering the drugs in a compulsive fashion, we increase the levels of dopamine D2 receptors, do we affect the self-administration of the drug? And that's what we did. And the study was led by Dr. Peter Thanos and this was the first study of this that he published, then doing others of this type. And this study . . . he was then a postdoc at Brookhaven National Laboratory, and he trained the Sprague-Dawley rats to self-administer alcohol. And when these animals were readily self-administering alcohol he then injected them with an adenovirus stereotactically into the nucleus accumbens. And in the adenovirus he basically inserted the dopamine D2 receptor gene. When you do that basically what results is a significant increase in receptors.

So this is the day of the injection of the vector. At day four receptors are 50 percent higher in the nucleus accumbens. The receptor overexpression is short-lasting, and that has to do with the fact that the vector was an adenovirus that does not integrate into the DNA, so its expression is very limited. And at day twenty, so the receptors go back to baseline around day ten. By day twenty he injected them a second time and then you can see it again. The receptors are increased again. And then he asked the question, In these animals in whom their dopamine D2 receptors are overexpressed in the nucleus accumbens, what happens to the self-administration of alcohol? And what happens is quite fascinating. It was dramatically reduced but it was not abolished. And you see it here, percent decrease in alcohol intake, approximately at four days, when you have the maximal increases in dopamine D2 receptors. It's almost 70 percent lower than when it was at baseline. It does not abolish it, it dramatically reduces it. And by day ten and twelve it's gone back to baseline.

When he injected the adenovirus again, again alcohol drinking behavior went dramatically down. And those were animals who were injected with the adenovirus, also into the nucleus accumbens, but the adenovirus did not carry the D2 receptor gene, and this was important to do because you wanted to be sure that the changes in behavior were not a function of inflammation from the adenovirus. And it basically shows that without the gene it does not have an effect.

So this particular study shows that indeed, in animals, overexpression of dopamine D2 receptors dramatically reduces alcohol intake, and does provide evidence that high levels of dopamine D2 receptors indeed may be protective, not just in general of not taking drugs, but protecting you against taking high doses of a drug, which is ultimately what leads to addiction—high doses of the drug—because we all control—more or less, if we're not alcoholics—how much alcohol we take. And so you see, these animals . . . it's not that the alcohol has been inhibited. What's also fascinating, and the study just came out this month, is that he's taken this particular study and now replicated it not in Sprague-Dawley but in animals that are
genetically inbred for their propensity to self-administer alcohol. And these are the preferring alcoholic rats. These rats rapidly start drinking alcohol and they prefer it over other substances. So in these animals we don't know which are the genes involved in these behaviors but they are genetically inbred for these behaviors, so there are genes linked with this propensity to self-administer alcohol. In these animals increasing dopamine D2 receptors also significantly reduced the alcohol intake, indicating something that to me is quite fascinating, that certainly dopamine D2 receptors per se are not accounting to addiction, not at all, but they may be modulating and regulating your propensity to become addicted or severely addicted depending on whether you have the genes or you have the environmental interventions that may then lead to the addictive process. And that's how I basically right now view the process of the dopamine D2 receptors.

**Brain Function and Dopamine D2 Receptor Levels**

Now the question is why is that so? I was speaking about sort of hypothesizing that what's happening is that [if] you have low levels of dopamine D2 receptors this will lead to . . . then lead to changes in the signaling to the circuits that are then responsible for motivation, for drive, as well as other cognitive operations. So we decided, as I said, in these subjects in whom we have also measured dopamine D2 receptors we can also [measure] brain glucose metabolism. And we had done that to actually address the question, are the changes in dopamine D2 receptors in any way reflected with brain glucose metabolism?, which as I say is an indicator of brain function.

So in the next slide what you're going to see first is . . . now you're not going to see images of dopamine D2 receptors but brain glucose metabolic activity in controls—three different levels of the brain. By the way these are basically . . . this is the lower level, one of the lowest levels of the frontal cortex, the area we call the orbital frontal cortex. And these are sequential planes, going from lower to up. And this is the scale. And this is a normal control, this is a cocaine abuser's. And what we found is cocaine abusers have significant disruption, particularly in frontal cortical regions. And of them two of the most affected regions are the orbital frontal cortex and the cingulate gyrus. And moreover, abnormalities in the orbital frontal cortex and anterior cingulate gyrus are not specific neither for cocaine, are also observed in marijuana addiction, are also observed in alcoholism, and in methamphetamine addiction. And these are just the values for the group, so they're significantly lower.

And so then the question was, Are these changes in metabolism in any way associated with the changes that we saw in dopamine D2 receptors? And the answer is yes, indeed. And this is the slide showing that we basically scanned the whole brain, and then measure metabolism in 33 brain regions, and then identified those brain regions for which the correlation between D2 receptors and metabolism was significant. And that was basically mainly in cortical projections of the dopamine system. And the strongest correlations were at cingulate gyrus and orbital frontal cortex.
And this is the individual values, measures of metabolism in the orbital frontal cortex, measures of dopamine D2 receptors. And you see it is the individuals with low levels of dopamine D2 receptors in whom we are seeing the decreases in brain glucose metabolism into the orbital frontal cortex.

And the same thing . . . this is a study that we replicated in methamphetamine abusers . . . the same finding. The subjects with low levels of dopamine D2 receptors are the ones that have decreased metabolism in the orbital frontal cortex.

And this was quite a fascinating finding because it took us by surprise. And it took us by surprise because in general most of the studies in drug addiction have concentrated mainly in limbic brain areas, and the frontal cortex was not considered to be important in the process of addiction. And what the imaging data basically threw at us was that one of the main areas disrupted with chronic use of drugs are the frontal cortical areas, namely orbital frontal cortex and anterior cingulate gyrus.

**Saliency and the Orbital Frontal Cortex**

Now why was this? This was perplexing. As I say, there was no data. There was one preclinical study showing involvement of the anterior cingulate gyrus. So it was perplexing because the frontal cortex had not driven the attention of people involved with substance abuse. However, the orbital frontal cortex had attracted the attention of those individuals investigating food-rewarding processes. And what these investigators had clearly shown was that the orbital frontal cortex is an extremely important area of the brain in signaling the saliency value of a particular stimulus. And they also showed [that] not only [is] the orbital frontal cortex an extremely important area in assigning saliency value to a stimulus, but also in changing that value as a function of a context.

Now why is that important? This is extremely important because when you have a reinforcer, a natural reinforcer, the value of that reinforcer is going to be affected by the context. And this was demonstrated very elegantly by Schultz, who, actually using electrophysiological technologies, was recording the cells in the orbital frontal cortex on primates. When you show a primate a piece of lettuce, it's salient and the orbital frontal cortex signals. But when you show that same animal the same piece of lettuce by the side of an apple, the lettuce no longer leads to firing of cells in the orbital frontal cortex, whereas the apple does. And he postulates that the reason why the lettuce is no longer able to activate the orbital frontal cortex is that the saliency value of this reinforcer when the apple is by the side is basically decreased. So that's the whole concept of the value of something is relative to something else. And that is driven by the orbital frontal cortex.
Moreover, what experiments had shown also in animals is when you destroy this area of the brain, when you damage it, something fascinating happens. When you take a stimulus you can actually make an animal press a lever for something that is reinforcing, so the animal presses a lever and basically gets food and presses the lever. But then the investigator removes the food and what the animal does is he stops pressing the lever. However, if the investigator damages the orbital frontal cortex something fascinating happens. The animal learns very rapidly to press the lever, just as if he were intact, or she were intact, so he presses the lever. But if the investigator now removes the food the animal continues to press the lever again and again and again, even though that lever is no longer reinforcing. In other words, the animal has lost its ability to change the value of what originally was a salient event. And that continuing to press the lever again and again when there's no longer a reinforcement, indeed, is very reminiscent to what many of my patients tell me. They say, "Doc, I don't even know why I take the drug. It's no longer pleasurable. I just cannot stop taking it," indicating that indeed, perhaps the role of the involvement of the dopaminergic system is by disregulation of these areas of the brain that basically are crucial in driving our behaviors. Because this is the area of the brain that would signal ultimately when something is salient and behaviorally relevant for the animal to take into action. And the other area, the anterior cingulate gyrus, is an area of the brain that in the other side of it allows us to exert inhibitory control. So something is very salient, we want to do it, but we catch ourselves saying, "Uh uh, this is not a good idea." It is the anterior cingulate gyrus which will be able to stop us from doing it. So what we're postulating is that in drug addiction, disruption of the dopaminergic pathways leads to disregulation of these frontal areas that are key both in motivating our behavior and two, in allowing us to exert inhibitory control.

And let me just go through these ones very rapidly, because it's late in the day, and just end up with my last slide in terms of what have we learned with the imaging in terms of what the role of the dopamine system is? The dopamine system is not the only neurotransmitter involved in drug addiction, it's not at all. But what these studies have actually been showing us is that in fact disruption of the dopaminergic system is evident in a wide variety of drug addictions and that this disruption in turn is producing dysfunction in areas of the brain that are involved with motivation and drive, inhibitory control, also very likely reward, nucleus accumbens ventral tegmental area. And an area that I have not even had a chance to go into, for which the dopamine system is extremely important, is also in facilitating memories, both by conditioned responses through the amygdala, as well as normal memory process through the hippocampus. So what it brings forth to light is that the process of addiction is not producing a disruption just in one brain region or circuit, but that it implies disruption of multiple circuits that are accounting for that very malicious phenotype that leads an individual to take that drug at the expense of his own incarceration, at the expense of many times losing the family, at the expense of basically losing the job, even when that individual says that the drug is no longer pleasurable. So this is the circuitry that imaging
actually has documented, not just through these studies but others, to be involved in drug addiction.

Now how do we take this [inaudible]? Well my perspective is, if we know that these circuits are disruptions, this is important because it is telling us that when we are designing treatment we should start to basically develop treatment that will strengthen motivational drive circuits that would interfere with conditioned responses such that the individual will not have this intense drive to take the drug when exposed to it, and finally that can promote plasticity of dysfunction of these brain circuits that have been either damaged by chronic use of drugs or, alternatively, may have been dysfunctional from the beginning and have made a person more vulnerable for taking drugs.

So my end slide is, of course, this research has been the effort of a wide group of very talented investigators at Brookhaven National Laboratory, which from this would have not been possible, and of course with the general support of NIDA and the Department of Energy.

And I thank you for your attention.

Questions and Answers

Richard Mayeux: We have time for a few questions.

Man: Allan [inaudible] at Columbia. Nora, that was a wonderful talk . . . may I also say stimulating talk. Two questions: One is, is there a relationship between the degree of drug abuse and dopamine availability? And the second question is, Have you looked at relatives of drug users to see if they also have lower D2 availability?

Nora Volkow: Yes, the question has to do . . . is there an association between, I guess, the severity of the history of drug use and the levels and the availability of dopamine D2 receptors? And the answer is there is a relationship with years of drug utilization. The problem of that relationship is confounded by the fact that the longer the years you've been taking the older you are, so there is the confound of the age effect. So it's not clear cut. We have not seen a significant correlation with the doses of drug reported by the abusers.

Have we looked at individuals' families? Yes we have. We have looked at . . . specifically for alcoholics and these are experiments that we're doing in collaboration with Henry [inaudible] and Bernice [inaudible]. And what we've seen is quite fascinating, so we're studying these children who have a father who's an alcoholic and another first-degree relative who's an alcoholic. And we measured dopamine D2 receptors. These individuals are adults, they are not addicted to the drug but have a strong, strong family history. And to my surprise and horror, because I was counting on this on getting a grant, I was expecting that the
receptors were going to be low. The receptors were significantly higher than normal controls. And of course that makes writing for a grant much harder.

But basically you have to look at the data and say well, what type of explanation you can make to that, and again that's basically based on our animal studies. We said well, perhaps this individual, even though they may have the genetics, are not alcoholics because for whatever reason [they] have the high level of receptors that are protecting them. And so this is . . . we put it in as a hypothesis. It was our pilot data for a grant. But then at [inaudible] I saw a fantastic study which is much more powerful than ours, even though it was only seven subjects. It was Mark [inaudible] and what he was showing was looking at identical twins discordant for alcoholism. And what he found was as follows: the alcoholic twin, as expected, low levels of dopamine D2 receptors. The discordant twin for alcoholism had significantly higher levels of receptors than normal control. So we found that in a much less powerful design, just looking at subjects, he is now finding this in identical twins, indicating that what we don't know is what is accounting for the increases in dopamine D2 receptors. The fact that his identical twins were discordant also in the levels of dopamine D2 receptors is telling you that the drive is not just genetic, that something, for whatever . . . we don't understand what it is. It's likely to have to do with issues such as stress from animal experiments. But again it's very tentative, what you can make of the explanation.

Man: What about nondrug addictions like gambling addictions and that sort of thing? There seems to be a sense, from what I'm gathering, that there's some sort of almost self-medication here to reach a desired level of stimulation. But have you looked at these kind of disorders?

Nora Volkow: We have. And I was intrigued very much because, I mean, my thinking has evolved, and like the previous speaker I've made some spectacular mistakes. So, but in a way one of the things that I was thinking . . . okay, low levels of dopamine D2 receptors, because that's what I started with thinking makes you more vulnerable. Now I think that high levels are protecting you. But so I was thinking that, that low levels made you more vulnerable. So I said well, is it just a function of chronic drug administration or do you see it in another type of condition where you have the same compulsive behavior? So we went for obesity. We went for obesity because it's a big-impact disorder. And what we found was actually quite fascinating, because it gave us some insight on a problem that has counted me . . . all of these have low levels of dopamine D2 receptors. What leads to the specificity of the drug as the reinforcer, right? And so we went to the obese people. These were morbidly obese patients—I think it was average 350 to 400 pounds . . . very difficult to do because they broke the camera. So we had to construct a new bed and after two years we succeeded. And what we found was the levels of receptor were significantly lower. But different from anything that we've ever seen in drug addiction was these individuals have increased hyperactivity in the areas of the brain, somatosensory areas of the parietal cortex, that are regulating perception of food stimuli. That was the lips, the tongue, the mouth; they were
hyperactive. And so what it gave me an insight . . . was in terms of why certain individuals with low levels of dopamine D2 receptors would go, for example, for a drug versus going for food? What it is keeping me, again, all of these are hypotheses based on data that you get . . . is that by having increased activity in these somatosensory areas it makes that person much more sensitive to the palatable aspects of food, which is one of the variables that would account [for the] hedonic properties of food. So it's more sensitive to that particular reinforcer so [inaudible] has low levels of receptors is going to be much less sensitive to other things. But the fact that it's hypersensitive to this one may drag him across this path. So that's where the insight came.

Now why is it that someone may, for example, take alcohol versus cocaine? There are individuals that take anything, but there are also individuals that will just stick to one drug. And you can start to understand that, that that's where the specifics of the genetics may come around. Alcoholic subjects, it's not infrequent that they will tell you, "Doc you may want to believe me or not, but if I don't drink I just feel perpetually anxious. So as a result of that I function much better drinking." So then you start to understand why would a person choose heroin as opposed to cocaine? It is likely that there may be there particularities in terms of disruptions of those systems that make you more vulnerable. And here I'm just speculating, but bringing up the notion that genetics may come from other neurotransmitters and what dopamine may be doing is basically either protecting you or alternatively making you more vulnerable.

Man: This is a follow-up to his question. I'd like to know references to that obesity research that you mentioned. And taking the three elements of your last slide, change, reward, motivational system, interfere with conditioned responses, and promote plasticity of dysfunctional brain circuits, how do you apply that to an addiction, say like a food addiction?

Nora Volkow: Well how do you apply that? First of all, one of the aspects, let's say conditioned responses, they are extremely important certainly in drug addiction, and it's also certain that they are extremely important with food. I find myself always eating a chocolate when I go and pay for something even though I'm not hungry because I see that chocolate. So you are conditioned to respond in certain ways. So now if you can control it more or less you're okay. But when it becomes pathological in my brain it would be wonderful if we could develop a medication that will interfere with that conditioned response. So you ask the question, what type of studies have given evidence that this is possible? So I went back and I said, do we have evidence that you can interfere with Pavlovian conditioned responses? And the answer is yes, investigators have been able to do it. There are not many studies, but the few studies there are, they basically show that drugs that enhance GABAergic activity interfere with conditioned Pavlovian responses. And one of the things that's fascinating right now in terms of promising medications for the treatment of drug addictions are drugs that are enhancing GABA, have actually
shown to be beneficial in alcohol in humans, and have shown in multiple preclinical studies to be beneficial in models of animal drug self-administration.

So that’s an example of plasticity, and that’s, of course . . . again I’m being very ambitious, but I think we should be ambitious now that we have the technology, the tools, and we’re understanding the brain. We know that for many years investigators on learning disabilities have used training in order to make kids with learning disabilities able to learn and read better. But what these investigators have now shown with images is that they can actually lead to some recovery of brain function, and if not the same area, they actually can make other areas take over for the function that account for the improvement in reading. So I'm putting forth the same concept and say why can we not then, using the same strategy, design behavioral interventions to strengthen, for example, the anterior cingulate gyrus, which is one of the areas that allows us to exert inhibitory control? And there are many ways that people are looking at that: you could do it with behavioral interventions, you can do it magnetic stimulation like it's being done now with patients with stroke, but we haven't yet applied this type of strategy for a problem like drug addiction. But I'm just sort of putting it . . . these are potential things for which we now have the tools to start to utilize.

Richard Mayeux: Thank you all for your attention this afternoon. This concludes the afternoon portion. And thank you to our speakers.
Brain and Mind
May 14, 2004

Nancy Kanwisher, PhD
fMRI Investigations of Human Extrastriate Cortex: People, Places, and Things

Introduction by David Cohen

I'm David Cohen and it's my pleasure to welcome you to the third and final session of the "Brain [and] Mind" symposium. Having been a university administrator for the past 17 years, I'm not at all sure that I still have the neuroscience bona fides to deserve a spot on this illustrious program, but I appreciably accepted the assignment as a generously nostalgic gesture that recognized the time some years ago when I worked for an honest living as a neuroscientist.

Now this session is entitled "Biology of Mind," and in recognition of Richard Axel's interesting metaphor yesterday morning, I'd like to take a moderator's prerogative and re-title the session as "The Ghostbusters." Now in this session, complex cognitive questions in the phenomenon of consciousness are addressed as biological-search challenges. And as Eric Kandel and I were commenting to each other yesterday, a session of this nature is something that we could have only fantasized about when we began in the brain sciences some forty years ago, yet the questions that will be discussed today are the kinds of questions that initially attracted many of us to the brain sciences, even though the prospects of any kind of experimental answers at the time were exceedingly dim.

That we're here today in a session of this sort really is testimony to the explosive development of brain research over the past four decades. Perhaps a surrogate measure of this has been the development of the Society for Neuroscience. In its first meeting 33 years ago, there were two to three hundred attendees. At its meeting this last year there were 30,000, and that is a growth over just the last 33 years of the existence of the society. Neuroscience is clearly one of the most vital and exciting fields in contemporary science, and that's a fact that even the physicists would concede.

Now we have four distinguished speakers in this session who continue the impressive array of contributors to this rather remarkable event. And I might just make an idle observation, if you look at the program and look at the degrees that are held by the participants in each session, it appears that if you were trained as an MD, you work on the brain; if you were trained as a PhD, you work on the mind.
And I'm wondering if there isn't some barrier in training of physicians that simply does not permit them to become ghostbusters.

Now the range in this session is extraordinary—it's from inferring the mechanisms of complex neurofunction from the activity of single neurons and functional neuroimaging, to biological, theoretical, and philosophical considerations of consciousness. It does indeed promise to be a provocative morning.

Our first speaker is Dr. Nancy Kanwisher, who's a professor in the Department of Brain and Cognitive Sciences at MIT and an investigator at the McGovern Institute for Brain Research. She received both her undergraduate and doctoral degrees from MIT, and after serving on the faculties of UCLA and at Harvard for some years returned to MIT to join its faculty in 1997. Now if you'll permit a brief apropos of nothing, whenever I hear MIT and Harvard concatenated as in Nancy's resume, it evokes a story I heard as an undergraduate at Harvard, and it's about a student who wields a cart full of groceries—this is in the Central Square Supermarket in Cambridge—he wields a fully loaded cart up to the register that says "Ten items or less," and the cashier looks at him in disgust and says, "Look, you're either from Harvard and you can't count or you're from MIT and you can't read." Since Dr. Kanwisher has been at both she can both count and read.

Nancy has been applying the noninvasive techniques of functional magnetic resonance imaging and magneto-electroencephalography, and if memory serves, I believe magneto-electroencephalography was developed at MIT some years ago, and I think it was by a namesake of mine, David Cohen, wasn't it? She's been contributing fascinating, compelling experimental results that identify and characterize regions of the human brain that mediate the visual perception of faces, bodies, places, and objects. She's been recognized by various awards including a MacArthur Foundation Fellowship, a Troland Research Award from the National Academy, and MacVicar Faculty Fellow Award from MIT. It's indeed a pleasure to welcome Dr. Kanwisher at Columbia, and I look forward to hearing about her fascinating findings on "fMRI Investigations of Human Extrastraite Cortex: People, Places and Things." Nancy.

MRI Studies of the Visual Cortex

**Nancy Kanwisher:** Thanks, David, and thanks so much to the organizers for inviting me. I've learned a lot and had a very good time so far.

So the question that's motivated my research for a long time is this, within a fraction of a second of viewing a completely novel and unpredictable image of a complex scene we human perceivers have already extracted the gist of the scene, and if we know the specific individuals or places or objects in that scene we've also identified those individuals. If that doesn't strike you as remarkable, reflect on this: Despite decades of effort, no current computer vision system is even close to what all of us can do in a fraction of a second. So the question is, How do we do it?
Now if I could answer that question I'd be as famous as Eric Kandel. I can't answer that question, I'm not that famous, but what I will do is dance around the edges of this question and show you the approach we've taken in my lab, which is this: Our general strategy is to try to get some clues into how vision works by looking at the functional organization of the parts of the brain that do it. So that's the stuff back here. Close to half the cerebral cortex is devoted to vision, construed broadly, including visually-guided action and thinking visually and things like that.

We are enormously visual animals. So how is all this expansive brain organized? Well we know a lot about the organization of visual cortex in macaques from decades of research where you can stick electrodes in and record from actual neurons, the gold standard in neuroscience. A great deal has been discovered, and so all of these visual areas up there have been characterized in great detail in monkeys, and even more impressively, the connectivity, the wiring diagram of the visual cortex has been characterized in great detail in macaques.

In contrast, until extremely recently, almost nothing was known about the organization of human visual cortex. So about ten years ago maybe two or three visual areas had been identified in the human brain. And then functional MRI came along. Functional MRI is just like regular MRI scans that you might get in a hospital for a knee or a kidney or a liver, except for two important things: One is that the images can be taken very quickly, upward of ten images per second, and the second is that the images are arranged in a special way such that they show, very indirectly but very well, a measure of neural activity. It works by way of blood flow. If a bunch of neurons in a focal region of cortex become active, that's metabolically expensive, more blood needs to flow to that region of the brain, and MRI looks at that change in blood flow. Also importantly, in contrast to other methods such a PET, which involve a radiation dose to the subject, as far as we know, functional MRI is completely safe. I certainly hope it's safe because I have very close to the world record of total amount of time spent inside these machines.

So the question that we've asked with functional MRI is very analogous to the question that Richard Axel talked about yesterday. He asked whether different parts of the olfactory bulb respond when animals perceive different odors. We ask whether different parts of the visual cortex become active when people look at different kinds of objects. So getting MRI machines to work is very fancy, I get zero credit for that, I had nothing to do with it. Using them to ask this question is really pretty simple. We stick subjects in the MRI scanner and we show them movies. So here's a diagram of the kind of movie that we show our subjects in one experiment. Time is going from left to right, each vertical stripe represents a period of 16 seconds during which the subject viewed a particular thing. So starting over here subjects viewed a fixation point for the first 16 seconds, then they saw a whole bunch of pictures of flowers flashing on the screen at a fairly rapid rate, then a whole bunch of animals, chairs, scenes, cars, faces, and so forth. We just tell the
subjects to hold still, look at the pictures, and while they're doing that we're scanning their brain.

**A Face-Selective Cortical Region**

We then take all the data from the functional brain scans conducted during this kind of experiment and we ask statistically whether there are any voxels in the brain that respond significantly more strongly during the periods when subjects are looking at one of these kinds of images, compared to when they're looking at all the others. The kind of thing we see is shown here. Here's one slice through a subject's brain, this is a horizontal slice, like this. Left and right are flipped here and on all my images, that's the radiologists' fault, they started that convention. And what you see over here is a little teeny region in this subject's right fusiform gyrus where the statistics are telling us that the signal is much stronger; that is, neural activity is stronger in that part of the brain when this subject was viewing faces than when they were viewing anything else. Now you shouldn't believe these nice activation images if that's all you see, there are a million ways to cheat and get nice images like this that are entirely spurious. The reason to believe this image is if people show you the raw data. So here's the raw data coming from that part of the brain during this scan. And you can see just eyeballing it that this little teeny bit of this person's brain produced a much stronger signal during the two periods when the faces were presented than during any of the other periods. You can see the signal's higher when they were looking at all these different kinds of objects than when they were staring at a dot, so it's not absolutely silent for nonfaces, but it's a whole lot stronger for faces than for anything else. So this is what we call a face-selective region of cortex in the human brain, and what does that mean? Well, it could mean all kinds of things. From the data I've shown you so far this could be just due to, for example, greater attention that people may pay to faces. We're social primates, we care about faces, maybe we just pay more attention to faces than anything else. It could have to do with simple low-level properties of the stimulus that haven't yet been controlled. There are any number of accounts of what this could mean, and you can't make a strong claim about face selectivity until you spend a long time testing all those other accounts.

So the way we've done that is what I call a region-of-interest approach. What we do is first identify that region individually in each subject with a scan much like the one I just described. We find the bits in that subject's brain that respond selectively to faces and we say okay, there's the face area, then we run a new experiment and we ask how that region responds under some new conditions. And we can quantify the magnitude of the response in that area to the new conditions of interest. And this solves a whole bunch of technical problems with functional imaging. One is that those face areas, although they can be found in virtually every normal subject, there's just a certain amount of anatomical variability, you don't know exactly where it's going to land in the fusiform gyrus. So this way we don't worry about that. We find it in each subject and we look there. It also gives us a huge increase in statistical power, and has other advantages.
So having done that for quite a few years on the fusiform face area (FFA), here's a summary slide of some of our results. So on the left there are four different subjects' face areas, here and—I can't see from this angle—here, and that's me, I have this posterior annex to my face area, and here's another subject. So you see there's some variability here. So over here you see in each panel an example of a kind of stimulus, and a magnitude of response. A 1- to 2-percent signal increase is a reasonably strong response for functional MRI, and what you see is lots of different kinds of face images produce a strong response in this region.

Let me just point out a few things. I don't know if you guys can see the pointer. Can you guys see this image? Raise your hand if you can tell what that image is of. Very few hands. Okay. Raise your hand if you can tell what this image is of from the top row in the middle. Okay. So you guys behave like our subjects do. Most of the time they see the face in this image. If you didn't see it there's nothing wrong with you, it's a little tricky. It's a profile of a face like this with the nose sticking out on the left and the eyes and the mouth. The interesting thing is this is the identical image, it's just upside down. So what this enables us to do is respond to all the kind of serious vision geeks who want to tell us your face area, it just likes edges like this or it likes contrast or it likes this spatial frequency. We can say no, all those things are pretty much identical in these two images, and here where subjects usually see a face we get a much stronger response than here when they don't usually see a face.

You may be wondering, why does this region respond to cat faces? I'm looking in human brains and we humans care about human faces. Well we wondered about that too, and then we thought well you know, cat faces look a fair amount like human faces, in fact some animal faces look very much like particular people. Okay, those were totally irrelevant, I just like to show them.

The FFA and Facial Recognition

Okay, so to get more serious about it, all I've shown you so far is that this little bit of cortex is pretty selective for faces. Doesn't tell us what it does with faces, that's what I'd really like to know. So what I'm going to do is describe one experiment done with Kalanit Grill-Spector, who's now a professor at Stanford. She was then a postdoc in my lab, and we asked, well, our overall strategy to try to figure out what this region does is to give subjects a difficult perceptual task in the scanner, make it really difficult so that subjects make some mistakes, and then collect MRI data while they're doing the task and bin our MRI data by what the subject reports, trial by trial. Then we can look for correlations between the magnitude of the brain signal in different parts of the brain and the subject's response in order to get a closer causal connection between the MRI data and behavior.

So here's the experiment that we did: We chose famous faces like Harrison Ford and various other actors and politicians that our subjects knew the faces of. We
presented them as a timeline from left to right, we presented them very briefly and followed them by visual junk, mask here, so that subjects would make some mistakes, and in this experiment we asked subjects for each face to give us one of three response options. Either it was the particular target face that they were looking for, in this case Harrison Ford, or it was some other guy, or it was nothing, just visual junk. So they have to make one of those three responses on each trial. We can then look at the magnitude of response in the fusiform face area on just the Harrison Ford trials as a function of what the subject tells us they see in those stimuli. So the stimuli are more or less the same, they're all pictures of Harrison Ford, and all that differs is what the subject reports.

So what you see here is—this is a timeline here—this shows you the typical MRI response. It's delayed, this peak is about six seconds after the onset of the stimulus, that's because blood flow regulation lags well after neural activity. But what you see here in the different height of the peaks is in red where the subjects correctly said yes, that's Harrison, in blue are the trials where they said some other guy, and in black are the trials where they said I don't see anything. Remember it's always Harrison Ford. So what we want to say is that this difference between the correct identification and the simple detection of faces, that reflects a correlation between this MRI signal and face identification.

This additional effect down here shows that this region is also correlated with simply detecting the presence of a face, even when you don't correctly identify it. And this pattern of response was shown in all the subjects here. Of course I'm subject five who has generally the best data, but everybody else showed the same thing.

So we want to use these data to argue that fusiform face area's involved in both detecting the presence of a face and identifying individual faces. But you might be wondering when you get a response correct, when you detect what you're looking for, maybe there's like a yippee! effect, maybe there's a party in your brain and you just kind of . . . everything turns on more. Also we want to know how specific is this response to faces?

So in another set of experiments we did the analogous experiment on a bunch of different kinds of objects, in this case guitars. So here subjects had to say is that an electric guitar, another kind of guitar, or nothing. We analyzed the data in the same way, and here are the results for the guitars, broken down by individual subjects. You can see there's really no correlation between the response in the fusiform face area and successful identification of guitars. We found the same thing for tasks on cars and flowers and so forth. So this shows the specificity of the fusiform face area in detecting and recognizing faces, but it's also of interest to ask about the regional specificity in the brain. What's the rest of the brain doing? Maybe the whole brain is more active when you detect or recognize a face.
So here I'm showing you from one subject, as the physiologists say a typical subject, most of our subjects look like this. This is a bottom view of the brain, and what's shown in black are the outlines of the fusiform face area defined in our usual way, and what's shown in yellow and orange are the regions that showed a significant correlation with face identification in that subject. And you can see there's a pretty good match. And there isn't much else, you know, spills out a little bit over the back right there, but mostly it's that region that's correlated with face identification, not the whole brain. And when we do the same kind of analysis on this same subject looking at other kinds of objects, you see that most of the regions that are correlated with correct identification of other kinds of objects are outside the fusiform face area. So that says that other objects are primarily categorized and identified using other bits of cortex, nearby but distinct from the fusiform face area.

Okay, so all of that suggests that the face area, what it's doing for us is both detecting and identifying faces, and that it plays little role in detecting or identifying other kinds of stimuli. But as Nora Volkow said yesterday, these kind of data are just associations, they don't really show a causal connection. A correlation with behavior is kind of as good as we can get with functional MRI, but the real test is if disruption of this region disrupts performance on the task.

So these are some data—not mine—this is from a paper by Wada and Yamamoto and they report the case of a gentleman who had an unusually small stroke in his fusiform gyrus. And I've outlined it here in the right, fusiform gyrus right there. This gentleman had a remarkably specific case of prosopagnosia, that's a severe deficit in recognizing faces. This guy can't recognize himself in the mirror, he can't recognize family members, he's absolutely unable to recognize faces, and yet interestingly, like many other cases that have been reported previously, he's very good, in fact they claim absolutely normal, at recognizing objects. Okay, so his deficit is very specific, and although I've been trying to get an MRI scan on him to make sure that his fusiform face area was, in fact, affected by this lesion. For various technical reasons we can't scan him, but I'll just show you my fusiform face area on analogous slices. What you see is, here's this guy's lesion and there's my FFA. They're really similar, so I think it's a good bet that this guy's stroke took out just his fusiform face area and produced a very selective deficit in face recognition. So that suggests that the FFA is not only correlated with face recognition but necessary for face recognition.

**Other Brain Areas in Facial Recognition**

Okay, so enough on the FFA. I'll just briefly mention two other regions we've discovered since then. I should say the FFA has been fun to work on but it wasn't exactly surprising given the prior literature on prosopagnosia, and lots of behavioral and physiological work all of which suggested special neural mechanisms for face recognition. So it's fun to go in and find it and characterize it, but it wasn't a big surprise.
In contrast, the next area that we discovered in collaboration with Russell Epstein, who's now a faculty member at U Penn, we call the parahippocampal place area (PPA). It's about a centimeter anterior in the brain and a little medial, toward the middle of the brain from the face area, and it's bilateral. You can see it here in four subjects, right there, these two little blobs. It is just behind the hippocampus but not overlapping with the hippocampus in parahippocampal cortex, and here's what it does: It responds very strongly when you look at pictures of places. It can be outdoor scenes or indoor scenes, it doesn't matter if there's any stuff in those scenes or if it's just a bare spatial layout. These pictures are really boring to look at when you're in the scanner, but the PPA is very excited about these stimuli. And when you take them apart the visual properties are very . . . here you can still see the spatial layout and the response is the same. And here the spatial layout information is disrupted even though the visual properties are very similar. Signal drops. Can't reach it with the pointer, but the abstract scenes made out of LEGO on the far right there produce a very strong response in this region there to objects made out of LEGO. So this leaves completely open the question of what this region is contributing to place perception, but it does provide preliminary evidence that this region's very selective for place perception.

A little sidebar. It's of interest both to identify and characterize the function of these regions, but it's also a lot of fun to use them to address other questions in visual cognition. So just to mention a few of the kinds of fun and games we've had with the FFA and the PPA, we've shown that the activity in these regions is very strongly modulated by visual attention. So if you have exactly the same retinal information hitting your eyes from a stimulus, if you choose to pay more attention to faces than places you can essentially crank up and down the dials on your own visual system. You can control the operation of your own visual system by visual attention. We can see this reflected in the magnitude of response in those areas, which is modulated by visual attention. Second, we can show using a bunch of perceptual tricks. In this case we used binocular rivalry that when exactly the same perceptual information hits your retina, if the stimulus is ambiguous such that sometimes you see a face and sometimes you see a place, we can see the activity toggling in lock step with the subject's reported percept in these two areas. These regions reflect not just what is striking the retina but what the subject perceives from that information.

And finally, we can remove ourselves from the stimulus altogether and just ask subjects to lie in the scanner, close their eyes and imagine a face or a place. And when we do this we see selective activation in the very same regions, in the face area for imagining faces, and in the place area for imagining places.

Okay, so back to these selective regions. The third one, and the final one, I'll describe only briefly that we've discovered. This is work—with Paul Downing, who's here, now a faculty member at the University of Bangor, Wales—is what we call the extrastriate body area. This is pretty weird, but what can I say, the data
dragged us here kicking and screaming. And what the data show is a strongly selective response in every subject we scanned to bodies and body parts. So here it is in four subjects, these are now vertical slices like this. This activation on the lateral surface is right next to visual area MT that's interested in visual motion, for those of you who know it, but not overlapping with it. This region here responds much more strongly to all different kind of bodies and body parts, compared to lots of different kinds of control conditions. To get control of the low-level visual properties of the stimulus here we compared stick figures of bodies versus scrambled stick figures. Much higher response when you see a body in the stick figure than when you don't, and silhouettes showing bodies versus scrambled silhouettes. Also very similar visual properties, much stronger response when you see bodies than when you don't. All of this, of course, leaves wide open the question of what this region is doing with information about bodies. It may be used in recognizing individuals, it may be used in knowing where your own body parts are for better visually guided action, it may be used in social cognition, in understanding other people's actions. So all of this is wide open and under active exploration right now.

Experience and the Visual Cortex

Okay, so I've described these three category-selective regions of the brain, shown diagrammatically here. The body area is not out in space, it's just around on the upper surface of the brain where I can't depict it here. Now, these regions are found in pretty much the same rough location in essentially all normal subjects, so this is showing us something very basic about the functional organization of the human brain. Second of all, the selectivity of these regions is very strong. This is not like lots of brain imaging studies where you scan a large number of subjects and you just barely reach statistical significance with a tiny effect size. These are all enormous effect sizes. And my intuition is big effect sizes are telling you more important things. But of course it raises a huge number of questions, of which I'll just briefly describe and discuss two.

One is, where does all this functional organization come from in the brain? That's a very hard question to answer because—let's go back for a second—there are two very different kind of caricatured simple stories you could tell about the origins of these regions, and they're both quite plausible. One says look throughout primate evolution. If you had to choose three categories of objects that were probably really important for us to be able to perceive clearly, and that was probably faces, places, and bodies. That's one story, so maybe natural selection has produced special-purpose machinery that's hardwired into the brain, and maybe that's why those things land in systematic locations in the temporal lobe. But here's another equally plausible story: Each of us individuals looks at these stimuli very frequently in daily life throughout our lives, and we know that the cortex is very shaped by experience, as several of the preceding lecturers have described. And so maybe the visual cortex just kind of does statistics on the input, and says, "I'm seeing a lot of these and those and those, and let's allocate big chunks of cortex to the things
we see a lot.” So those are both plausible stories. I would love to figure out which of those or what kind of compromise position is correct for the face, place, and body area, but because they're both plausible here, and because I'm working with humans, I don't know how to answer that question.

So I'll address a different question, which is closely related, and that is whether experience alone is ever sufficient for the construction of a category-selective region of the cortex. This is work with Chris Baker on the right and Jia Liu. So here's how we tested this: We used the test case of visually presented words. Why words? Well, our experience with words is very high, especially for geeks like my students and me. Our experience with words is pretty close to our experience with faces, although importantly, it does start later in development. In contrast, unlike faces, people have only been reading for a few thousand years, and this is not usually thought to be enough for natural selection to produce special purpose machinery. So what that means is that we can use words as a kind of test case. If we find regions of cortex that respond selectively to visually presented words, there'll be a kind of existence proof that extensive experience can produce a selective region of cortex.

So lots of people have reported the existence of a visual word form area, and there's a lot of work on this, but the critical questions haven't really been adequately answered. One is, How selective is that region? And the other is, Is it really resulting from experience? So I'm going to briefly describe an experiment that Chris and Jia and I have done. We scanned subjects while they looked at words and pictures and lots of other things, and our first question was, Can we find a region in their brains that responds selectively, or that is significantly much more strongly to words than pictures? So here is such a region in one subject. It's small but we see this in most subjects in the left fusiform gyrus, and it's producing a much stronger response to words than pictures. I told you before not to believe activation maps, so what we do here is replicate that with a further study, where we make sure we get the same result in the same subjects in the same scanning session, and we throw in some other conditions to see how selective this region is. So here first is the replication of the higher response to words than line drawings. Well what about other stimuli? This is the response to Chinese characters. Our subjects don't speak or read Chinese. Here's a response to strings of digits. So it was quite selective. Strings of digits are a lot like letter strings. But here is this response to letter strings. So already we've learned one important thing: This is not, as many have claimed, a visual word form area. It's maybe a letter area or a letter string area. But for the purposes of testing the effects of experience that's okay, I'll take letters, same deal. So this region is quite selective.

However, I haven't really shown you yet that this is due to experience. You could say well, maybe there's a whole kind of space of different feature selectivity in the cortex, and whatever features happen to be present in the words but not the digits and character strings maybe they would be there even if you never learned to read. It's hard for me to find subjects in the Cambridge area who don't know how to
read English words; however, we can move to a different language. So what we did was we looked at the response to people looking at both English words and Hebrew words as a function of whether they were readers of Hebrew or nonreaders of Hebrew. So first these are the data from these subjects showing you the response in that region to English words versus pictures in our non-Hebrew readers, and English words versus pictures in our Hebrew readers, who also read English. Okay, so that just shows you the kind of thing I already showed you. Now the critical question is what is the response to Hebrew words in these subjects? Okay, so here are the non-Hebrew readers. You can see that the response to Hebrew words is quite high, quite a bit higher than to pictures; however, it's significantly lower than the response to English words. What about the Hebrew readers? It's a whole lot higher. So I should say we only have three subjects here so far, so we need to run a few more subjects. But it's so clear in these three subjects that I'm already reasonably confident that this result will stick. And what it seems to be telling us is that the difference in the response to Hebrew in this case, in this case shows us a very strong effect of experience on the selectivity of this region.

So in answer to the first question—How do these regions arise?—well, what we could say is at least some of the selectivity in this general part of the brain is based strongly on experience. It doesn't tell us that that's the origin of the face area, the place area, and the body area, but it's a kind of existence proof that says that might be the origin.

**Selectivity in High-Level Cognition**

Okay so the final question I'll address only very briefly, and that is whether this kind of high degree of functional specificity that we see in the face, place, and body areas is really something about the visual system. That it's very kind of modular, it parcelates the problem of vision into all these separate little functions allocated to different regions of cortex. And maybe that's special about vision.

Or, do we ever find that kind of selectivity for very high-level cognitive functions? And what I'll say is I think the answer to this question, this is very much an empirical question, I think you have to test each case. So, for example, there are a lot of very interesting claims about a region in the parietal lobe up here in the left hemisphere that's been argued to be selectively involved in understanding number, that's a very high-level function, and there's quite a bit of evidence for that. However, in my lab, what we show is we can find that region, it is involved in understanding number, we can replicate that. But it's also involved in lots of other things, like visual attention, like selecting responses that are appropriate for the visual stimulus, like lots of other functions. So I think the answer in many cases will be no, there aren't highly selected bits of brain allocated to high-level cognitive functions, there are general purpose bits that do a lot of cognition for us. And by the way, that's also true in vision. I focused on the category-selective areas but
there are also what appear to be very general-purpose areas that represent the shape of more or less any old object, as far as we can tell.

So some things are very general like this, but I'll show you just very briefly one region that is remarkably specific, beyond my wildest expectations. So this remarkable and brilliant student arrived in my lab four years ago, Rebecca Saxe, and she said, "I want to study theory of mind." This is the topic that Mike Rutter mentioned yesterday. And she said she wanted to do this with functional MRI. I said, "Well, that's lovely, but one, I don't know anything about theory of mind; and two, I don't believe for a moment that it's localized anywhere in the brain. And if it's not, then functional MRI is the wrong tool to use." But luckily she was undaunted and she spent three years in a very focused fashion approaching this question. So she ran a version of a Sally-Ann task that Mike Rutter described yesterday, in which subjects read a little verbal description of a person who's presented with some information about another person and what they know, and they are then asked a question that requires them to distinguish between what that person knows about the world and what is true about the world. The way you pull those apart is to make the belief false. Otherwise, you can answer the question by reporting on what's true of the world. So she used the Sally-Ann task, and she showed that this region in the temporal parietal junction is very strongly activated when subjects perform those false-belief tasks. Several other people had shown related things before; however, what they hadn't done is get obsessive, like we like to in my lab, and really address each of the alternative accounts we can think of. So Rebecca did this and she ruled out every plausible alternative explanation I can think of, a few which are shown here. One is this region does not simply get engaged whenever you do identical problems about nonmental representations. So instead of doing Sally-Ann task, instead of having Sally watch Ann and having the subject report on what Sally knows, instead of Sally we have a Polaroid photograph. So now we have instead of a mental representation, a physical representation. So you have the logically identical question, you ask the subject in the scanner about what's in the photograph when it differs from reality. No dice, this part of the brain is not interested. So it's not just responding to any representations that differ from reality. It's not simply false representations. So if you ask people questions about true beliefs you get the same high response in that region. If you ask people to reason about hidden causes, you describe physical events for which they have to infer some hidden cause, you don't get much of a response in this region. And if you ask people questions about people that don't refer to their beliefs, like about their physical appearance or other properties, or their actions, you don't get the strong response here. So despite my extreme skepticism before I went into this, I'm now convinced that this region is very selectively involved in representing the contents of other people's beliefs. And that's a very high-level cognitive function. So I'd say yes, at least some very high-level functions do get small regions of cortex allocated to them.

And I'll stop there. Thank you.
Questions and Answers

I guess I can take maybe one question. Yes, back there.

[Question inaudible.]

We've tested that. We had what we call—okay, the question was, Is there a snake area? And I have to apologize, I tried to get on the Internet last night to show you a snake picture but the Internet connection in the hotel wasn't working, so sorry. We did test that condition. In fact I skipped over this because I didn't have time. But we tested pretty much every other visual category we could think of for which we could make an even remotely plausible story that there might be a special-purpose bit of cortex. And that included one condition where we had both spiders and snakes. Paul Downing, who did the work with me, is like, not phobic about snakes but he is about spiders, and I'm the opposite, so we threw them together in one category. We couldn't find any selective responses. Of course, importantly, the fact that we don't see them doesn't mean they're not there. It's the easiest thing in the world with functional MRI to not detect a selective response that's actually in there for any number of reasons, including for example, the neurons that are selective may be interleaved with other neurons below the resolution of your MRI scan. But the main story is we tested lots of other conditions and we didn't see anything with the kind of size and robustness of selectivity that we see in each subject for faces, places, and bodies.

[Question inaudible.]

I haven't, but it's . . . okay, so the question is, What happens in autism subjects who generally don't like to look at faces, what are their responses? I haven't studied this, but Mike Rutter mentioned yesterday a couple of studies of people who have looked at it. The reports in the literature, there are two or three studies that I know of, argue that you do not get a fusiform face area-selective response to faces, and instead nearby regions of the cortex that respond more strongly to objects seem to be engaged when autistic subjects look at faces. That's suggestive, but I'm very cautious about this for two reasons: One is I know of at least one other lab, I think two, that is getting different results, and I want to see what happens with that. And the other is there are many ways in which that experiment could fail to find the face selective response. For example, the very fact that the autistic subjects don't like to look at faces may mean that in the scanner they're not even [inaudible] the faces. I don't think the published studies tracked eye movements to make sure that the subjects were looking at the faces. This is critical. Even if they're looking at the faces, you need to make sure they're attending to the faces, so they need to be given a task that requires them to perceive the faces. And until all that is done and replicated I'm a little bit skeptical.

Man: Since science is a progression of errors, does the work, in terms of organization, relate back to the work of Lashly and [inaudible] and others who have
done work in the area to vindicate some of their concepts of fifty years ago? And
the second thing, have you done any work in the area of auditory perception—
music, for example? There's an area of the brain with perfect pitch and so on,
which you can perhaps do some specific work. I wonder if you can elucidate that.
Thank you.

Nancy Kanwisher: So the questions were, first, How does this relate to the
historical work of Lashly and others? The second question was, What about music
and auditory perception?

On the first question, yes absolutely, I'm addressing the oldest question in
neuroscience, which is the pendulum on this topic. The fashions on this question
have swung back and forth reliably every few decades for a very long time, and
that question is, To what degree is function localized in the brain? And I showed
you some functions that are very localized in the brain, but I also noted that many
others are not. I gave the example of understanding number, but there are many
other examples of things that are not localized in the way that I described. So yes
I'm still pursuing the same old question. I'm sure that neuroscience will still be at
this for a long time. And I should say that my position on this is a little bit
idiosyncratic. Many of my colleagues who run very similar experiments would
stand up here and tell you that what the data show is massive overlap in those
functions, so there's not any agreement on this right now.

Second question on music and auditory perception—I haven't studied this but
there are a number of interesting studies. The closest one to my work is a study by
Pascal Belin that was published in Science in 2001. It actually came out of a
conversation that Pascal and I had where he was studying tones and auditory
perception. I said, "Well, what would be the auditory equivalent of the face area?"
Well easy, it would be a voice area, right? Not speech because that's different, but
voices, because you can recognize individuals from voices, you can . . . like faces
you can pick up emotions and all kinds of other information, age, about the
speaker. And sure enough he went and did a number of studies and found in the
superior temporal sulcus, up closer to auditory cortex, a region that's quite
selectively responsive to human voices, and this included stuff like laughing and
crying and sighing, it did not include stuff like speech, because, that is, it did not
require speech, the nonspeech sounds also activated it. But it wasn't just any
human sound, so when they had people clapping or walking or making other
sounds that were clearly human did not activate the area, it was specifically voices.
And there's lots of other work on music, but I don't know that literature that well. I'm
probably supposed to stop.
Brain and Mind
May 14, 2004

William T. Newsome, PhD
Decision Making and the Neural Representation of Value

Introduction by David Cohen

David Cohen: Well that was a spellbinder. Nancy, you'll have to come back and update us annually. I think this is a comment the way many of us have felt almost every talk in this symposium.

Our next speaker is Dr. William Newsome. He's a professor in the Department of Neurobiology at Stanford and a Howard Hughes Institute investigator at Stanford. He received his undergraduate degree from Stetson University and his PhD from the California Institute of Technology.

He began his faculty career at the SUNY, Stonybrook, in the Department of Neurobiology and Behavior, and at this point I think it's probably fitting that I shift from calling him Dr. Newsome to Bill, since I was the chair who recruited him to Stonybrook for his first position. Unfortunately the department lost him in 1988 when he went to Stanford, and I think it was in no small measure because Bill and his wife, Zondra, never could acquire Long Island as a second language.

Bill has been particularly courageous in tackling especially challenging research problems, challenging both conceptually and technically, problems in visual perception and cognition. He's been extraordinarily innovative in devising behavioral paradigms which combined with single-cell recording to address questions about deep levels of visual perception, decision making, and the role of value in decision making. He's appropriately collected the expected array of awards, including the Rank Prize in Optoelectronics, the Spencer Award from Columbia University for highly original contributions to neurobiology, [and] the Distinguished Contribution Award from the American Psychological Association. Most recently he's been invited to deliver the 13th annual Marr Lecture at Cambridge University, and the King Solomon Lectures at Hebrew University in Jerusalem. And I think you're on your way to that, aren't you, Bill? Bill was elected to membership in the National Academy in 2000. And it's with some nostalgia and some pride in my recruiting taste that I'm indeed pleased to be able to present Bill Newsome who will speak on "Decision Making and the Neural Representation of Value."
Internal Representation of Value

William Newsome: Thank you, David, for the introduction and for that first job, which of course is extremely important.

So I'm going to be talking today about decision making and the neural representation of value, and many of you have seen this title and wondered what the hell is he going to talk about? And one day neuroscience, hopefully, will get to the point where we can talk about ethical values and higher values of humans and how we arrive at those things, but I'll be talking about something more mundane today, which is value in the economic term. Okay? So we're all seeking constantly to maximize return on our investments, right? So many of you had a decision to make a few moments ago, which was whether to stay here and listen to this talk or whether to go outside and enjoy the morning. And you did some calculations about value and potential return on your investment in time, and fortunately for me most of you decided to stay. And I'm hoping that I won't disappoint you in that decision. Some people, however, resonated with David Cohen's lovely Freudian slip this morning when he talked about this "oppressive lineup of speakers" for the symposium, and they're out enjoying the morning. And my best wishes are with them.

I want to say before I start that all of the work that I'll talk about today is a collaborative venture between myself and two extremely talented graduate students in my lab, Leo Sugrue and Greg Corrado, so when I say we, those are the guys that I'm talking about in addition to myself.

Okay, so let's talk a little bit about this value thing and how we got into it. My lab, as David alluded to, we've studied visual perception, have fifteen, twenty years on the books studying visual perception, which led us naturally to a study of decision mechanisms that we've been working on, simple forms of decision making in primates we've been working on for the last—really intensively—for the last five to seven years. And this little diagram here shows you the classic sensory physiologist's view of decision making, and I am a sensory physiologist. So there's a world out there that provides sensory input to the brain, and there are sensory systems in the brain that analyze that world—in Richard Axel's phrase yesterday "deconstructs that world"—and loads a representation in sensory areas of the cortex, and then ultimately at some higher level of the cortex, decisions have to be made about what's out there: Is it predator, is it prey? And of course those decisions that get made influence your motor-output structures. If it's predator, you're likely to run or fight; if it's prey, you're likely to get ready to pursue and hopefully have some hunger satisfied or something like that. So in this sensory physiologist's view of the world, decision mechanisms follow directly upon the sensory input. But economists and psychologists have known for a long, long time that this is a highly impoverished view of decision making, and that there are internal representations within the brain of value, the likely value, of certain outcomes and certain actions that we all take, and these internal representations of
value influence decisions just as much as does the sensory input, and sometimes more so.

So one of my favorite examples like this is a trout fisherman going out to fish in the stream early in the morning, top of the morning. He's got a decision to make: which pool do I cast my fly into, which rock do I stand on? And the sensory evidence here is really rather weak because, after all, he can't see the fish, they're down there under the surface of the water, right? So how does he decide? And the answer is he decides because he has some internal representation of value. Where on this stream this time of morning at this time of year have I had luck in the past? In other words, he has a reward history of trial and error and where he's been rewarded or failed to have been rewarded, and that history that all of us are carrying around with us all the time influences these internal representations of value which in turn influence the decisions that we make.

Now, for a person who studies the brain if you're interested in decisions, it's obviously extremely important to bring this under control and start being able to see physiologically how these internal representations of value work and how they influence decisions. But this is a difficult thing to do in a laboratory, this is rather hairy. If I'm a sensory physiologist and I'm studying vision, I have a place to stand. I stand on the sensory input. I can present a particular visual stimulus to a human subject or to an animal subject time after time after time, the way that Nancy just described for you, and I can see what areas of the brain or what cells in the brain have activity that's correlated with that stimulus. If I'm a motor physiologist, I can train a human subject or an animal subject to reproducibly output motor behavior over and over again, and I can see what sorts of activity in the brain correlates with that motor output. But how in the world are we going to bring this under control, the sort of stimulus, action and value?

The Matching Paradigm

Okay, so that's the first question that we've got to answer here. And the short answer is that we are not going to have to reinvent this wheel, that there is a group of psychologists who've worked long and hard in the 1950s and 1960s, Richard Herrnstein and his colleagues, to establish what they call the matching law as a way to assess an animal's evaluation of stimuli. So the notion of the matching law that Herrnstein worked . . . he worked largely with pigeons, okay, and the pigeons had a choice of which lever to peck, lever A or lever B, so they had a choice, it was a free-choice paradigm. And the law says that animals allocate time or responses among competing behaviors in proportion to their relative frequency of reinforcement. So you can put this little equation down which just says the response on lever 1 divided by the total responses—so that's just the proportion of responses on lever 1—is equal to the proportion of the time that the animal gets rewarded for pressing lever 1. Now the good news here is that this reward fraction is under experimental control. Herrnstein could control this, right? So he can program his computer to decide how likely the animal is to get rewarded when he
presses lever 1 versus how likely he is it get rewarded from pressing lever 2. And here he's manipulating the percentage of reinforcements on key A, all the way from zero, which means that key B gets rewarded all the time, up to one hundred, which means that key A gets rewarded all the time. And this shows you data from three pigeons—how frequently the pigeons responded on key A—and you can see that there’s this nice linear relationship—these pigeons are obeying the matching law very nicely so that when they get rewarded infrequently on key A, they choose key A infrequently. And when they get rewarded frequently they choose key A frequently.

Now Herrnstein’s brilliant insight here was that this kind of behavior . . . if an animal actually matches, now we have a behavioral handle on value because we can use the responses, the proportion of responses, that the animal makes, that's under his control, that's under the animal's control, and we can measure it, and we can use that as an index, a proxy if you will, for these internal valuations—how the animal's valuing the likely outcome of these two choices.

So we have adopted this matching paradigm to use in our laboratory, and I'm just going to give you here over the next half hour an insight into the way that systems neuroscientists like myself think about these problems, and the way that we try to approach these problems in the laboratory. Now I'm going to be talking about this value thing, but the basic plot is the same whether we're talking about visual perception, attention, memory, decision making. If you get the gist, the hang of the way we think and do things here, you can then apply this to all sorts of other sort of cognitive situations.

The first thing is that work in my field always starts with behavior. I've already outlined for you behavior, a central behavioral question. So we believe in this field that the brain's primary function is to produce intelligent and interesting behavior, okay? So our questions always start with behavior, and I've already posed for you a behavioral question about value. And we are going to talk now when we discuss this matching paradigm about how we measure value rigorously at the behavioral level. Then we also bring theory to bear. And we in the theory, we're going to say can we formally describe—and by formally I'm actually talking about mathematically—describe how value is computed within the brain; i.e., can we build a model? We're going to try to build a model that actually duplicates the animal's choice-making behavior. And then finally we're going to address some physiological questions, we're going to say what are the neural circuits within the brain that implement this computation?

And this is relatively new work in my lab. The first paper has just been accepted for publication, actually, it's not yet published but will be very soon, and so you should regard this as a progress report. These questions are not completely answered at this point in time, but again, David, have me back in five years and we'll see what happens.
Matching Experiments in Rhesus Monkeys

Okay. One little aside before I go forward. All of the experiments that I'm going to talk to you about today are done in rhesus monkeys—awake, behaving monkeys—who actually perform a task and we measure their brain activity while they're performing the task. Now there are some kinds of neurobiological questions that are very important, such as the one Rod MacKinnon talked about yesterday, that you don't actually have to have an animal on line to do. You can work in tissue culture, you can work all sorts of different reduced preparations. But those of us who study the nervous system and want to understand how the nervous system produces intelligent behavior have to use an awake, behaving nervous system while it's actually in the process of producing behavior, right? There's only a limited amount you could learn about a car engine without the car engine actually in action and doing its thing. If the car's just sitting there turned off you can make all kinds of theories, but how do you test them? Well you've got to make manipulations and crank that car up and see if it works. Well it's the same thing in studying the nervous system, and we actually have to work in a functioning nervous system.

Now there are some things you can do in simpler nervous systems. As Eric Kandel showed us yesterday this inelegant little animal, *Aplysia californica*, can actually perform remarkable memory feats, and Eric has been enormously successful in unraveling some of the circuitry underlying memory in *Aplysia californica*. People work in rodents frequently because rodents are good spatial navigators and they're good olfactory discriminators. So there are some things that different organisms are good at. But by far and away the most cognitively versatile animals that we can actually bring into the laboratory and do experiments with are monkeys, which is why we work with our rhesus monkeys. And I'll have more to say about that in just a moment.

Okay. So our first task here is behavior. We're going to the matching law, we're going to train monkeys on a matching task, an ocular-motor matching task, so they're going to use their eyes to do this matching thing, and that's deliberately chosen because the ocular motor system, the motor system that controls the eyes, is the best-known motor system in the primate brain, so we can take advantage of lots of background work. Okay, now I'm going to describe the task for you right now, and you have got to hang with me here, okay, because if you don't understand this task the rest of the time you'll have higher value out with a cup of coffee and enjoying the morning, okay? So follow this task for me now.

So an animal sitting in the primate chair, and the animal comes out and does this every morning so he's utterly accustomed to this sort of thing, and he looks at a TV screen in front of him, and we have him playing this little economic matching game. An animal sitting in the midst of an apparatus called the scleral search coil apparatus that allows us to measure his eye position extremely accurately at any instant in time. So our computer knows from millisecond to millisecond exactly where this animal's pointing his eyes. The first thing that happens is the little
fixation point comes on the TV screen, and the monkey has to look at it, and that lasts for about 300 milliseconds. After 300 milliseconds two targets come on, a red target and a green target, and ultimately the monkey's going to move his eyes to one of these two targets. But first of all there's this little delay period. The red and the green come on and there's a delay period that's variable in length from one to two seconds. So one thousand one, one thousand two, and after that delay period the fixation point dims and the monkey then moves his eyes to one of these targets or the other target and he either gets rewarded or more often he does not get rewarded, and I'll tell you about that in just a moment, and then he holds his eyes on that target that he has selected for two to four hundred milliseconds, and then the fixation point turns bright again and he moves his eyes from the target back to the fixation point, okay, so that's finished with one trial, he's made one choice. And then these two red and green targets may be rerandomized spatially, if you watch carefully they're going to switch positions, there they go, and then we just come back to this point here, okay, and we start all over again. So he has to wait one to two seconds, he makes a choice, and so on and so on, and he just goes around in this circle making eye movement after eye movement, choosing the red target or the green target. Okay?

Now the critical thing is, How does he get rewarded? So the monkeys are thirsty when they come out to work in the morning, and they'll sit here and work for this sort of thing because they're going to get a shot of juice, or sometimes it's Kool-Aid, we find what they like the best, and they'll sit here and do this thing for a couple or three hours while we acquire behavioral data and electrophysiological data, so they're highly motivated to get these rewards, and they'll sit here and go in this loop, making eye movement after eye movement. Now here's the critical thing, here's where the economic and the value and the matching comes in, is, How do they get rewarded? All right. Now what you have to do, you have to imagine that there's a little stopwatch attached to each one of these two targets; there's a red watch attached to the red target and a green one attached to the green target, and these guys are counting down, and on average the red one's going to count down to zero once every 10 seconds, and on average the green one's going to count down to zero once every 30 seconds. And when those clocks count down a reward becomes available. And the very next eye movement that the animal makes to that color he picks up a reward. All right.

Now it's important to realize that these clocks are independent. They're not synchronized to each other at all; it's independent things going on on the two colors, and the times, the actual times, are a Poisson process—that means they're random. And this is extremely important. What that means is that this clock counts down on average every 10 seconds, but on one trial it might be 2 seconds, the next trial it might be 21 seconds, the next trial it might be 13 seconds, then 7 seconds, but on average it's going to go down every 10 seconds. So on average rewards become available on this target three times more quickly than they become available on this target, but there's a lot of noise, a lot of stochastic
random noise here, and in order to estimate the rates you've got to integrate across that noise over time, okay?

Now when a reward becomes available, when the clock counts down to zero, and the reward's there it stays there, it hangs out right there on the target, until the monkey comes and collects it. And that's very important because it means that it's logical for him to collect occasionally on the lean target. So this one that's counting down rapidly, you can call that the rich target, the one that's counting down less rapidly we can call that the lean target, and you might say well why would he ever go to the lean target when the chances are always higher that he's going to get rewarded on the rich target? And the answer is that these probabilities are cumulative, so after you've gone to the rich one several times this guy up here has a very high chance of being rewarded, in fact it'll have a higher chance eventually than the red one will because the probabilities have been accumulating there. So it pays off economically; you get more water, you get more juice, for the return on your investment if you include this guy in your plan and you go out and check him out occasionally. Okay?

So you can imagine yourself sitting here doing this, and you're just sitting here like the monkey and you're going to the red one and the green one. Most trials you don't get rewarded, but you pick up a reward once every third trial or so on average, and your job is to figure out which one is the rich one. That's called the exploration problem in foraging theory. But then you also, once you figure it out, you want to exploit your gain. So it's the exploitation-exploration problem here. And it turns out that matching—when animals match—it's an optimal solution to that problem, okay, which Herrnstein figured out thirty or forty years ago.

So this is—I don't—it's hard to have sort of question-and-answer interaction. This is a time when I would stop and ask you, but is this reasonably clear here? Okay, all right, I see some heads nodding which is a good sign. All right.

Now, so the monkey is matching if his response ratio equals the reward ratio. So in this example where we've got ten seconds and thirty seconds he's matching if he makes 75 percent of his choices to the red one and 25 percent of his choices to the green one.

Now here's the final trick that we play. We have frequent unsignaled changes in the reward ratio, so the monkey's cruising along, he's got the world by the tail, he's judged that this one's three times more valuable than this one, and then suddenly the world changes. And we might reverse them, okay, or we might make them fifty-fifty, or we might make them 6-to-1, something like that, and so the monkey's—and people, we've done this on people also—they've got to be careful because the world changes. And this is typical of the world you and I live in, right? I mean the world changes on you—things happen, so to speak. So the animal's got to watch for these unsignaled changes in the reward ratio. All right.
Data on Valuation in Monkeys

Now this is data, these are real behavioral data that are collected over about eight hundred trials, this is one of these sessions. And I'm going to step through this slide fairly slowly because this again is another critical slide. What we're plotting here is the monkey's cumulative choices toward the green target against his cumulative choices toward the red target, and we've got eight hundred trials' worth of choices here, and the monkey's choices are shown by this brown line. And I'm going to unpack that brown line for you in a moment, or this yellowish brown line. The purple lines show you when we changed the world on the animal, and I'll unpack that for you in a moment.

First of all the brown line. By following this brown line you can actually know exactly what each of the eight hundred choices was. So if the monkey only chose green, if he just chose green time after time after time after time, this brown line would come right out along this axis. If the monkey only chose red time after time after time after time, the brown line would go out along this axis. But he's switching back and forth, so the brown line traces out a trajectory through this space, right? And so every little rightward hitch means he chose green, and every little upward hitch means he chose red. So here he goes red, green, red, red, green, red, red, green, green, and so on. And here's he's making more red choices than green, here he's making more green choices than red, and that slope flattens out, okay?

Now these purple lines show you where we changed the game on the animal. So this first epoch of about eighty trials here he has a 1:1 ratio, he has an equal likelihood of getting rewarded for the two, okay, so the slope of this thing is 1, and you can see that this brown line is following the purple line nicely. Now right here there's an unannounced change and the rewards go 3:1 in favor of green. And you can see that very quickly this monkey starts making more green choices, and this brown line starts heading out to the right. At this point we go 3:1 in favor of red, and again he makes a change and the brown line starts heading more upward than rightward. And here we go back to 1-to-1, and there we have radical shift, 1-to-6, and here we have 6-to-1, and this brown line is nicely tracing the purple line which means simply that the animal's matching. This shows you that we've successfully trained this animal to do what Herrnstein said they would do to get an optimal solution to this exploration-exploitation problem, which is to match the ratio of their responses to the ratio of the rewards that they're experiencing. And remember, he's doing this with no cues to the fact that the world has changed, he's just got to figure out from his choices and his success or failure at getting rewarded, okay? All right, so the monkey's matching.

Now here's something extremely important. In the theory part of this talk, we are going to ask, What computation is the animal doing? What's the little computation going on in the monkey's mind to assess the value? And we can pick up a very important clue already. And the important clue is this: see how rapidly the behavior changes at the inflection points? It's as though this animal's extremely sensitive to
the change in the reward rates right here, even though this is stochastic and variable, and he changes his behavior almost as quickly as the actual reward conditions change. Now this one's a little more slowly, right, this is like a little motorboat spinning on its axis in a harbor, and this is more like a ship turning around this corner, but mostly these shifts are very, very quick, these inflection points are very quick. Now what that means is that the computation of value has to be local in time. If the animal were estimating over this entire period here, trying to get his estimate of the likelihood of getting a reward, he would not change this quickly, he'd keep cruising past this inflection point and he would only change very slowly out here, okay? So this animal in order to compute value must be paying attention to the most recent trials that he's experienced. Now he can't be paying attention just to the last trial, or else he'd be a slave to what happened on the last trial. He would never be able to estimate the probabilities, so he's got to estimate the probabilities integrated over a certain number of trials or not too many or else he wouldn't be able to change quickly and he'd lose rewards, he wouldn't be sensitive to the changing of the world. So that's an intuition that I'm trying to give you about this calculation, that it's local in time, that he's assessing something like the last few trials. And we're going to put quantitative description on that; we can really quantify that and put numbers on it, but that's a key intuition about this computation. Okay.

A Computer Model of Valuation

So up 'til now we've asked, How do we measure value rigorously at the behavioral level? And we've seen that following Herrnstein we can train monkeys to perform an ocular-motor matching task, and the monkey's valuation of the two targets is directly indicated by the way he apportions choices between the two alternatives. Now the next little thing is theory. Can we formally describe how value is computing within the brain? And, that is, can we build a model? And I've already given you now a key piece of insight that whatever model we build is probably going to be estimating value based on what has happened in the last handful of trials, okay.

Now here's the model that we've actually built, and this model we build and we implement it on a digital computer, okay, and it's a simple model of matching behavior, and what we're going to do is put this model together on the computer and then we're going to feed the computer the same kind of history that we fed the monkey and we're going to ask the computer to do the matching task. So the computer now is going to make decisions, we're going to feed it two choices, we're going to say red or green, computer, and the green's going to say well red. Okay? And we're going to reward the computer or not reward the computer according to the same algorithm that we've used for the monkeys, we're going to change the world on the computer just as we have changed it for the monkeys, and we're going to make a little model here of the computer's decision process, the way it computes value, we're going to make this explicit and mathematical, the way that it computes value, and ultimately we're going to see does the statistics of the
computer's choices match the statistics of the monkey's choices? And if we can reproduce the monkey's choices on the computer we have a good sense that we're getting our hands on the reality, the internal reality, that's going on inside this animal, and that of course is our goal here, to figure out what it is in the monkey's brain, how is the animal computing value, okay?

Now here's the model and it's very simple. It looks more complicated than it is. So imagine the computer sitting here making these choices, red, red, green, red, green, green, red, green, and at any point in time we can look back over our shoulder and see the previous fifteen trials or twenty trials or thirty trials of history, and the history is encoded in these ones and zeroes here. So a one means that one trial ago he selected red and got rewarded, so when there's a one here it means he got rewarded. A zero means that two trials ago he selected red and failed to get rewarded. Now these three blanks mean that on those three trials he chose green, so we go down here and we see he chose green and got rewarded, chose green and got rewarded, chose green and failed to get rewarded, there's red, okay, and so time goes from left to right, it means we went green, green, red, red, green, green, red, red, and zeroes and ones show whether he got rewarded. So we have two streams of history here, we have a stream of history on the red target and a stream of history on the green target, and what we're proposing is that the animal and our model integrates this reward history with what we call a leaky integrator, that means that he computes his income—how much luck I've had on red and how much luck I've had on green—and he discounts this the further it goes back into the past.

Now that's just a physical statement of the intuition that I showed you that the monkey has to be integrating locally, okay? So the further back something goes in time the more he tends to forget about it and says it's irrelevant to what I'm doing now, I want to pay attention to what's happened recently. And $tau(t)$ describes many trials back the animal's looking, okay? And in our model $tau$ is a free parameter, so when the computer's doing this and we're simulating all of this on the computer we can vary $tau$ around, okay, and see how the computer performs as we vary $tau$, and we're going to do that in just a moment.

So we're computing the income we got on red, and income just means, you know, these ones or the ones that are really going to count integrated by this sort of thing here, and we've got some measure now discounted in time of our income on red. We've got some measure discounted in time of our income on green. And we simply divide the red income by the total income and that shows the fractional value of red. Now if we had green in the numerator then we would have the fractional value of green. And the red is just one minus the green; they're directly related to each other. And this is what we call the fractional value, and it simply says what's the relative success, what's the relative income I've accrued on green versus red, or buying Dell versus HP, right, what's the relative success that I've had integrated over some short amount of time? And then whatever this fraction turns out to be we just say that's the probability with which the computer's going to
choose red, so we're postulating that's the probability with which the monkey's
going to choose red. And then once we have the probability, like it might be 0.7, so
if he's had good luck on red recently this fractional value might be 0.7, so that's
pretty high, 0.5 would be random, it means I don't know which one it's going to go,
and we set the probability equal to 0.7, and then we just flip a weighted coin that's
0.7 heads, 0.3 tails, and we flip the coin and that's the computer's choice. So then
we can do hundreds of thousands of experiments on this computer and when we
do this we actually see that the computer's behavior mimics the monkey's very
beautifully for a particular $\tau$ of eight trials, okay?

Now remember $\tau$'s our free parameter, and it says how far into the past are we
looking in order to make our judgments of income, and the answer is that we get a
beautiful match of behavior when $\tau$ equals eight, eight trials. And that comports
with our intuition about the monkeys, right, that they had to be doing this sort of
thing locally. So this just shows you that we can reproduce that behavior very
beautifully with a computer. And what this shows you is a systematic evaluation of
$\tau$. So what we're doing here is showing you the harvesting efficiency. Of all the
rewards that become available, how efficient is the computer at harvesting these
rewards, so this is harvesting efficiency as a function of $\tau$. So remember that's a
free parameter; this governs how far back into the past we're looking to judge
value, and you can see the intuition that I tried to give to you earlier, which is that if
you're looking way far back in the past, 250 trials back into past, your efficiency is
very low, because you're missing those turns. The world is changing on you and
you're not changing appropriately because you're averaging over a long period of
time, so your efficiency goes down. If you're only looking one or two trials back
your efficiency goes down because you're a slave to what happened on the last
trial. And where the computer is optimal is right here with a $\tau$ of about eight trials,
and it turns out that those two blue dots show you exactly where the data lie from
the two monkeys that we've trained on this task. Okay? And this is amazing, right?
The two monkeys have found this optimal place, okay, that the computer tells us
optimal is seven to eight trials ago, and that's exactly what the monkeys are doing.
And what's even more amazing here is that the monkeys are approaching the
optimal efficiency that the computer achieves. Now remember, these monkeys are
like you and me, okay, they're working but they get distracted, they don't get
distracted by the Web or something like that, but there are noises in the hallway
and they get distracted, you know, their leg itches and they get distracted, toward
the end of the day they're not very thirsty anymore and the juice doesn't matter so
much anymore and they get distracted, but nevertheless their harvesting efficiency
is very close to what our best computer model can get to, okay?

**Physiology of Value Computation**

Now this is a bit of a long-winded way of saying to you that I believe we've gotten
this value thing under control. We can implement this on a computer, we can
reproduce the animals' behavior, the animals' time window that they're looking over
appears to be almost exactly what the computer's time window that the computer
is looking over. So we posed this question for theory, we said, Can we formally describe how value is computed within the brain? And what we found is that under our conditions the value of past rewards is computed within a relatively short time window, weighted roughly exponentially, so there's that exponential decay, with a time constant of about seven to eight trials into the past. So now we actually have a description of the computation that's going on in the monkey's head, okay. Now this is a formal modeling mathematical description, but what we really want to know now is how does this get implemented in the brain? And that of course is a physiological question, that's what us neurobiologists ultimately are all about, what are the neural circuits within the brain that actually implement the computation?

All right, so I'm going to show you a taste of the physiology that we've done. But first of all let's go back to our model and just remember now that the key parameter in this model is fractional value, okay? This is what we're now postulating is the internal variable. Remember I told you that as a sensory physiologist you stand on the sensory stimulus. As a motor physiologist you can stand on the movement that the system puts out, but when you're looking at these internal cognitive variables that we're postulating where do you stand? And where we're going to stand right now is on fractional value because we know that this kind of calculation that yields fractional value can model the animal's performance extremely well. So this is our postulated internal variable, and we can go through as the monkey performs the matching task and we can calculate fractional value on every single trial that the monkey performs because he, too, has a history, and we can just apply these filters, we know the right tau now, it's about eight trials, we know the right tau, and we can apply these filters, and we can calculate this putative internal variable on every trial. And the question is, Do neurons in the monkey's brain seem to code this putative internal variable? And I wouldn't be up here if it didn't, okay? I'd be giving another talk.

Okay, so let me tell you a little bit about the macaque brain. This is the lateral view of the hemisphere, this is the front of the brain, this is the back of the brain. There are a lot of things we know. This is the visual cortex that Nancy was talking about back here. There are three areas that are high-level suspects for us to begin recording from, and these areas are high-level suspects because we know that they're involved in the high-level planning of eye movements, and they are the frontal eye fields up here in the frontal lobe, the lateral intraparietal area in the parietal lobe that I'm going to actually show you data from, and an area down here in the midbrain called the superior colliculus. And we have only recorded from LIP at this point, and I'll show you some data from LIP, but ultimately we're going to record from these other areas as well.

This is a coronal section through the monkey brain; it's a slice like this, and this is LIP buried in the sulcus right here, and there's one on each side of the brain, and I'll tell you a little bit more about that in a moment. This is just to remind you that if we look at the cortex blown up and stained for cell bodies, every little purple dot here is a cell body, this is the top of the cortex, this is where the gray matter meets...
the white matter, and the cortex is a laminated structure and it's made up of hundreds of thousands of cells under any square millimeter of cortex, and the physiologists' stock and trade is to get these microelectrodes into the cortex. These microelectrodes are etched metal. The tip of the electrode cannot be seen with the bare eye, you have to look under a microscope. It's insulated down to within ten microns of the tip, and if you get this bare metal tip up close enough to a cortical neuron you can record the action potentials from single cortical neurons and study them one neuron at a time. This is what Nancy Kanwisher referred to as the gold standard for electrophysiology, really look at the action potential traffic of individual neurons. Now these electrodes don't hurt when they're in the brain, there are thousands of humans running around with these electrodes in their brain right now, as Gerry Fischbach told you yesterday about this patient with Parkinson's disease, they have these electrodes in for deep brain stimulation for therapeutic purposes, and they don't hurt because there's no pain receptors in the brain. You can do surgery on conscious human brains, you can take a knife and slice the cortex. You wouldn't want to do that, but you can do it and the patient won't say that hurts, they might say hey, I can't see anymore, what happened? But they're not going to say that hurts, okay, so it doesn't hurt. All right.

Now one other thing to tell you about LIP, LIP has been implicated in the spatial control of attention, and also eye movements as I've suggested to you. And one thing you need to know, because it's critical for what I'm about to show you, is that LIP on the left side of the brain codes the right half of space, and LIP on the right side of the brain codes the left half of space, so they're going to be operating in tandem with each other, sort of push-pull, left half of space versus right half of space. So Leo, my graduate student, gets his microelectrode into LIP, we know which LIP we're on so we know in this particular experiment we're on the left side of the brain, the electrode's on the left side of the brain, so we know that this LIP is coding the right half of space, which is shown by this little yellow egg yolk here, and we locate that and then we have the monkey perform the matching task. And one of these targets is in the part of space that this LIP cell cares about, and the other target is on the opposite half of space, so it's in the other LIP, but we're only recording from one of these guys at a time.

So the monkey's playing the matching game, and he's got two choices. He can move his eyes to the target that's in LIP's response field, okay, in the half of space it cares about, or it can move its eyes to the other target that's out of the response field. There's only two choices, right, because of the way we've set this thing up he could go into the response field or he can go out of the response field. So this is critically important, notice now the visual situation is the same here, red target to the left, green target to the right, although that flips on some of the trials, but in this particular situation the visual environment is the same, and within each of these conditions the motor output is the same. Here the eyes go right, here the eyes go left, okay? We've only got two conditions. But within all of those conditions now fractional value is varying, right? So sometimes he's going to choose this green target and because of his recent history, which we can calculate, right, we've got
our model now, we can calculate fractional value, and sometimes he's going to choose this one saying, "Oh boy, oh boy, I've got a good shot at getting a reward here," and sometimes he's going to choose this one thinking this is the lean target, but I need to go check it out every once in a while, and I don't have such a good shot at it, okay? So fractional value is varying even though the motor act is staying the same. Same deal here, he looks out sometimes and fractional value is skewing all over the place, okay? But we've got motor controlled, we've got sensory controlled, and the question is do the neurons track fractional value? And I'm going to show you some neural activity now, and this just shows you—remember, our plot here, fixation point comes on, monkey looks at the fixation point, and then these two targets come on, and the critical period is right here, after the two targets come on and this one- to two-second delay period occurs before he makes his eye movement, and he's valuing these two targets and he's trying to decide which one am I going to go to, and the question is, What does the neuron do during that interval when he's trying to decide?

And now I'm showing you what one LIP neuron does. This is activity, this is the neuron's activity measured with that microelectrode and it's plotted in spikes per second, so this is action potentials per second, as a function of time. This is the time when those two targets first come on, then we have this delay period, and zero over here is the time when the monkey actually makes the eye movement. Here he's going into the response field, and you can see there's this big burst of activity when the two targets come on, and then we settle down to a firing rate of around thirty to forty spikes per second until the eye movement gets made. Now the trick here is that he's gone to the right, but what we've done I've segregated out all the trials where he has a high fractional value, this is the top third of the trials with the highest fractional values where he's saying all right, I'm going there and I have a good shot at getting rewarded.

Now here is the activity on the lowest third of fractional values. This is where he makes the same eye movement, but he says I don't know, I don't have a very good chance here, and you can see the activity is lower. And then there's an intermediate third, and in the intermediate third the activity's actually intermediate throughout the trial, okay. So this is one condition where he chooses the target inside the response field. Now the other condition is where he chooses the target outside the response field, and when he goes outside the response field, now remember we're recording from this neuron and he's choosing the other one, and for high fractional values of this red target he's saying all right, I'm going out and I've got a good shot at getting rewarded, we have very low activity. When he's going out but he has a poor chance of getting rewarded, we have a higher level of activity, and so on intermediate. So this neuron's activity is nicely modulated by the fractional value. And we can take a bunch of neurons like this and average them together, and the picture cleans up nicely. So you can see that here we got a lot of activity when the monkey's choosing the target in the response field, as we would expect. We've got low activity when he's choosing outside the response field. And
this activity titrates out nicely by value. So the blue is 0 to 20, 20 to 40, 40 to 60, 60 to 80, 80 to 100. Okay?

So the bottom line here is that in this very first exploration what are the neural circuits within the brain that implement the computation, the bottom line here is that fractional value is represented in LIP, and the effects are often as large as those of the eye movement itself. Now there's one important thing for you to look at here, and that's dynamics. What's interesting, one of the things that's very interesting here is that time zero, remember time zero is when the two targets come on, and for the first 200 or so milliseconds of this response there is no effective value. All these curves are tracking together and value only hits LIP at this point here, two or three hundred milliseconds into the response. Now that's interesting, right, because that means that value is probably not calculated in LIP. Remember fractional value is integrated over the last seven or eight trials. By definition value has to be persist over trials, but there's nothing persisting right here. There is no value signal present in LIP when those two targets first come on, there's nothing persisting. What that means to us is that fractional value is, probably the original calculation is done somewhere else in the brain, and then it's getting downloaded to LIP on every trial. And one of the 60,000-dollar question is, Where is that? So the dynamics imply that the original computation of value is not performed in LIP itself.

Future Research

So the things we want to know in the future, and I told you we're just beginning to study, the things we want to know in the future is: Where and how is that color-specific representation of value computed? So remember color is the key to value here; it's the red target or the green target that have high values, so wherever it's being computed it's got to be color tagged. And then the value gets computed and presumably downloaded into LIP, and we want to know how does that color-specific representation become transformed in the spatial representation in LIP?

So with just the last couple of minutes here I want to return, take up a gauntlet that Richard threw down yesterday. Richard Axel, if you'll remember, talked to us about this odor responses, and the fact that when you present a particular odor to an insect you get activity at multiple locations in the glomerular map and in the mushroom body. And different odors will give rise to different patterns of activity, so the code for individual odors is not a matter of one neuron, one odor, but there's a distributed code, a distributed network that's very complicated, and the question that Richard raised was, Who's inside the brain reading that code? How do you look at a pattern of neural activity that's complex and across many structures and actually make a decision about what olfactant is actually out there in the world? And this is something just by asking that question that way, who's reading the code, it's the fallacy of the homunculus; it's like there's a little person, a little man inside the animal who's reading out the code. And somehow it's not a little man inside there that's doing that but the brain itself is doing it. And Richard said that
the ghostbusters would come today and try to decide for you, or for us, what it is or who it is that's reading out the code.

Now I don't have an answer to this question, but I have a way of thinking about it that I like better. If you really think about it, how do we know if a *Drosophila* can actually distinguish between two olfactants? Well you can only know that if there's a behavioral test, right? If you can devise a test or devise some natural behavior for the animal where the animal does one thing under olfactant A and he does another thing under olfactant B. In other words, the final common pathways that ultimately reveal to us the animal's perceptual ability is his muscles, right, it's what he does, it's where he goes. In our animals ultimately our revelation of value, this internal representation of value, is the animal's patterns of eye movements from one trial to the other. Now none of us think[s] that cognition really happens in muscles, okay, but there are areas of the brain that we know control motor systems, so I've already told you that these areas, LIP, frontal eye fields, and superior colliculus, we know on other grounds are involved in the high-level control of eye movements. And so to me it makes the most sense thinking these are the areas of the brain that have to program up the next movement, prepare it, they have to select the movement and program the movement, and that whole process of selecting the movement depends on reading out those codes about sensory information and about all these internal variables. So I personally think that we will find more and more in the coming decades that much of what we think of as high-level cognition is really sorted out in what we have traditionally thought of as premotor areas, okay? And I don't know exactly how that's done, what the algorithms are, but I suspect that the answers are going to lie in these kinds of areas.

So that's a progress report where we are to date. I think we're just at the edge of—those of you who are interested in psychology and economy will say there's a whole slew of questions here you can ask, you know, What about utility? Have you thought about the animal's utility, how utility changes with motivation? This is going to be a very exciting area of study, I think, and again, David, in five years maybe we can come back and have another look at it.

Thank you very much.

**Question and Answer**

Okay, so David gives permission for one question. There's a man at the microphone here.

**Man:** It seems that in this room we here have opened a tremendous can of worms, maybe not equaled since the first splitting of the atom, if I'm not exaggerating. If we're not careful it may be followed by a Manhattan Project, but it might be in the business school or political science or psychology. I think I've heard on [the] news that fMRI is being used for advertising, may be used by political groups. So do you
think this is a real danger now? And do you think that perhaps it’s time for the scientists, before it gets out of hand, to speak up on controlling this technology?

**William Newsome:** So that’s a very reasonable question, and certainly I think that all of neuroscience more and more will have impacts on society, and that there will certainly be ethical and societal issues that are going to be raised because of advances in neuroscience. I’m not very worried about the whole fMRI advertising thing. The one thing you should realize is that in my line of work the behavioral science precedes the physiological science, okay? It’s the people working on behavior who have the insights. Herrnstein had this insight about value long before we got into this thing, and we’re just using Herrnstein’s insights to try to work out the mechanisms in the brain that underlie this sort of behavior. Similarly in color vision, back in the nineteenth century we knew all about trichromatic vision, and it was only in the middle of the twentieth century when physiology caught up. What I’m trying to say to you is that in terms of advertising the people on Madison Avenue know a helluva lot more than any scientist knows right now. They’re the behavioral experts in this. Those marketing people, I mean all of politics these days and all of advertising is an exercise in human perception. And they know what appeals to humans and what makes humans lean this way or that way with their choices. And physiologists are just coming along afterwards and explaining, you know, why it is that this happens and where this comes from in the brain. Now can it ultimately make advertising more effective? Maybe so, I don’t know, I don’t know. But I don’t see that as nearly as much of an ethical issue at this point as, for example, you know, all of the issues about neural transplants into human brains and things like that, you know, to what extent do we use human subjects inevitably as experimental guinea pigs versus the great therapeutic benefits that can come? And I see big ethical challenges in the area of understanding behavior. As soon as we start understanding behavior mechanistically then do we just turn behavior into a machine, and how does that affect our legal system? If my brain just made me do it, if all my choices I make because my brain made me do it, how can I be held legally responsible? But this is not an unprecedented choice, right? We’ve dealt with diminished capacity and we’ve dealt with things like this in jurisprudence for decades and centuries now. So I don’t see this as fundamentally new. There will be issues that are raised, but at this moment I don’t see anything terribly pressing. I mean I’m not getting ready to go and outfit all of your brains with electrodes so I can stimulate you and control your thoughts. Okay.

**David Cohen:** Thanks very much, Bill.

We’re going to take a 15-minute break. We are running behind so we’ll hold it to 15. I’ve just been asked to inform you that the organizers and speakers have put together an online bibliography and reference guide for the symposium. And at some point on the screen we’ll give you the Web address.
Brain and Mind
May 14, 2004
Session III: Biology of Mind

Christof Koch, PhD
Towards the Neuronal Basis of Consciousness

Introduction by David Cohen

David Cohen: Our third speaker this morning is Dr. Christof Koch, who holds the Lois and Victor Troendle professorship of cognitive and behavioral biology at the California Institute of Technology. He is also professor of computation in neural systems at Caltech. Born in the Midwest, Christof had a peripatetic youth, lived in Holland, Germany, Canada, Morocco. It was in Morocco that he graduated from the Lycée Descartes in—I'm sorry—he studied physics and philosophy at the University of Tübingen in Germany, where he earned his PhD in 1982, and after four years on the faculty at MIT, he moved to Caltech where he has remained.

Christof has published prolifically, both articles in books on the neuronal basis of visual perception, attention, and consciousness, and has had a long-standing interaction and collaboration with Francis Crick on the dialogue about consciousness. He's boldly engaged in an experimental program to study consciousness, the topic of his latest book, The Quest for Consciousness: A Neurobiological Approach. The book was reviewed in one of the issues of Science last month and I'll offer you the final sentence, "The Quest for Consciousness is a brave attempt to fuse the best of scientific thinking with one of the central aspects of human existence." The title of his talk this morning is "Towards the Neuronal Basis of Consciousness," and he will describe a two-pronged approach involving psychophysical and fMRI studies of humans, and animal studies that combine behavioral methods with single-cell recording and pharmacological and genetic interventions. It is my great pleasure to present to you Dr. Koch.

Christof Koch: Thank you very much. Actually, I cut half my talk yesterday night, so I'm just going to talk about some electrophysiology and some behavioral experiments in mice. Most of the time we are all conscious, hopefully you're still conscious, you haven't gone to sleep yet, and so what . . . I mean the state of being conscious, of being conscious of my voice, of being conscious of colors, or being conscious of pain or pleasure, of being angry or being you, those are all simple aspects of our existence. And by and large science, for practical and methodological reasons, has not considered those, even though they are so central to our—I mean they're natural phenomena, they do seem to occur, and we'd like to understand the scientific basis of these.
Now my talk's divided in three parts, so the first part is some conceptual work, some conceptual [inaudible] work that Francis Crick and I have done over the past twenty years, to construct a framework for how we'd like to think about the neural basis of consciousness. So as John Searle is going to tell you in the next talk, consciousness or the mind-body problem at large—it's one of the oldest problems in physics or metaphysics or philosophy, and of course the ancient Greeks have had a lot to say about it. At the heart of that problem is the problem of qualia, which is, how is it that a physical system like my brain or your brain undoubtedly, or the brain of a monkey or [the] brain of a fly undoubtedly are . . . how is it that some physical systems at certain times of their life have these subjective states? Not always. Not when I'm in deep sleep, not when I'm under anesthesia or not when I'm dead, presumably, and not all of—I mean not all complex systems have those states. My gut doesn't have them, although it's very complicated. I have an enteric nerve system down here in my gut, contains between 50 and 150 million neurons, and there's very little evidence that by itself is conscious, or many other systems both in nature and in our artifacts that we construct are presumably not conscious, so what is it about a subset of them that gives rise to these subjective feelings?

And this has been hotly debated among thinkers and philosophers and scientists in particular for the past twenty years. We think the most central aspect of that, what philosophers call qualia, which are the elements of consciousness—the red of red, and the painfulness of pain, philosophers refer to those as qualia—how do qualia rise out of the firing, out of electrical activity, synaptic activity, metabolic activity in the brain? For now, we prefer to leave that problem aside, for tactical reasons. We like to concentrate on the problem of the neuronal correlates of consciousness. By the correlates we mean, What are the minimal neuronal mechanisms in your head, or in your body in general? What are the minimal neuronal mechanisms that are jointly sufficient for any one specific percept, for any one specific conscious percept?

Now, Francis Crick and I focus, again for purely tactical reasons, for now we focus on sensory perception, on conscious sensory perception just because it's very easy to manipulate, in particular, vision. Visual psychologists have discovered over the last hundred years, and have perfected a whole range of techniques that allow us to systematically manipulate what you see and what's physically present. So effectively we can now do what magicians do all the time. Magicians in front of your eyes they distract you, they have a beautiful bikini-clad assistant next to them, and they distract you by using these well-honed techniques of attention distraction. They distract you and so they can make things disappear, although physically you're looking at it. And I'll show you an illusion like that, which we can also do, so we can now begin to manipulate the relationship between the physical stimulus that's on your retina in the case of vision, and the subjective percept in your head. And we can do that much better for vision than for other modalities, and we can do that much better for sensory modalities than we can do that, for example, for self-consciousness, which to many people is sort of central to consciousness. It's something very difficult to study. You can study it in humans but it's difficult. It's
now being made possible with fMRI, but it's even more difficult to study self-consciousness in animals, which ultimately is where we have the best source of knowledge about the neural basis. So that's why we prefer to study visual consciousness. Although the belief is that consciousness is a feature of biological systems selected by natural selection over some, you know, tens if not fifty or a hundred million years. So it's likely that the central aspect of visual consciousness brings the system some advantage, some evolutionary advantage, that's also accorded to other aspects of consciousness, like consciousness for emotion or self-consciousness.

**Neuronal Correlates of Consciousness**

So once again the focus is on the experimental work, and there's a great deal of experimental work now that focuses on What are the minimal neuronal mechanisms that are sufficient for any one percept?

Our thinking sort of has evolved based on thinking of others, for example, physiologists like Bob Desimone and John Duncan, or going back earlier times to Hobbes, the idea is that your brain contains on the order of 20 to 50 billion neurons, and these neurons, they compete heavily against each other. They compete against each other, in other words, they can't all simultaneously be active for all sorts of reasons, partly because, of course, your brain would go into epileptic seizures, which obviously has to be avoided, and you have excitation inhibition among neurons, so you have coalitions of neurons that sort of . . . you can think of the brain a little bit like a Christmas tree, you have 20 billion or 50 billion little electrical bulbs on this and they all flash. These are the action potentials that Bill was just talking about in his talk. And so they all flash at different times, and they compete against each other, in other words, you have one group of neurons that suppresses other neurons. And ultimately consciousness arises out of the interaction among these groups of neurons. There's going to be one or several coalitions of neurons that correspond or that are jointly sufficient, that are sufficient for any one given percept.

One characteristic of conscious perception, as remarked upon already by William James at length, which is the fact that you typically can only be conscious of one or a few things at a time. The brain actively . . . there seem to be mechanisms in the brain that prevent two very similar things from being the focus of attention at any given point in time, you can only be conscious of one thing, and then of course you can rapidly switch your attention and your consciousness to something else. And so that ultimately, that's expressed by neurons that compete for each other. And so the trouble is our tools are very imperfect, we have tools, we have wonderful tools like Bill was telling you, electrophysiology, where we can listen to 1 or 2 or 50, or now with advanced technology, micromachining, we can listen maybe to 50 or 100 neurons, but we are sampling from a hundred neurons out of a sea of 20 or 50 billion, and so that's a practical problem that we're facing.
And so the belief—you have this coalition of neurons, and ultimately the winning coalition for any given point in time. And the winner of that coalition is . . . the representational content of the winner of that coalition of neurons is what you are conscious of. So it can be a voice and then it rapidly switches to an image and then it rapidly switches to the fact that you know your leg is itching.

**Zombie Systems**

As we know from our own personal experience, and this is, of course, a point that was much remarked upon by Sigmund Freud: Much of what goes on in our head bypasses consciousness; much of what goes on in our life we do totally automatically. And so there's this whole set of systems that Crick and I call Zombie Systems. These are highly attuned sensory motor systems that by the time you learn them effortlessly you do them automatically without having to think about them. You do them mindlessly in the sense that you don't have to think about them. So this includes, you know, in the morning you get up, you tie your shoes, you drive to work, and you type on the computer. Anything, sort of things like driving, like playing basketball, like climbing, like dancing, all these activities like moving your eyes, like moving your limbs, like reaching out and grabbing something, all of the things we know from experimental psychology and from clinical studies, are automatic. We do them at a very high level of proficiency; in fact, that's the point of training, that's why you train and train and train, so you don't have to think about them. And, in fact, very often if you do these activities so well, typically if you think about them, if you stop and think about [them] consciously it'll interfere with your performance. And evidence seems to be that the sensory motor system that is highly trained up, like including eye movements that typically don't have access to working memory, that if you use a system that requires working memory you then invoke a second set of systems, a system that seems to correlate with consciousness. So the claim is that you have this architecture where you have these two systems, on the one hand you have all these automatic sensory motor systems that control most of your life, and then you have this additional system that's much more powerful that allows you to do any arbitrary complicated task. And this is the system that empirically seems to be associated with consciousness. This system also has access to planning; in fact, that seems to be one of its key characteristics. That if you want to do planning, if you want to think, you know, suddenly there's a fire here, how do I get out? You know, how do I leave this building? I then have to bring to memory, I have to recall where's the entrance, how I can get to that entrance, that's a system that involves consciousness. And so the function of this system, Crick and I believe, is to plan. That's one of the key functions of consciousness, and that's probably one of the key reasons why it arose during evolution, to enable the system, to enable the animal, to do planning, to do things beyond the stereotypical response that the animal had learned.

This hypothesis has some anatomical correlates, in particular since we are arguing that consciousness is principally involved in planning, therefore we surmise that
the neurons that underlie the consciousness, or the NCC, the neural correlates of consciousness, that they have to have direct access to the planning stages of the brain, which by and large are in the frontal part of the brain. If you look at the neural anatomy based on a monkey, we know scandalously little about the detailed neural anatomy in humans. Based on the neural anatomy in the monkey, if you look at one of the best explored areas in cortex bar none, which is the visual cortex that Nancy showed you already, it’s at the back of your head. In fact, you can feel there’s a little bump here at the back of your head, and your visual cortex is a little bit above it. We know that because if you get hit on the head there you may see sparks or flashes of light. That’s what happened to cartoon characters in any case when they get hit there. So that’s your visual cortex. This is your primary visual cortex, the first entry point of the visual output from the eye through the thalamus into your cortex. This particular area is also called V1. This area does not have any neurons that project into the forward part of the brain; in particular, it doesn’t have any neurons that directly project into the planning stages of the brain. Therefore, we surmised ten years ago that neurons in V1 are not sufficient for visual consciousness, that they may be important for seeing, just like my retina is important. Clearly if I don’t have my eyes anymore I cannot do normal seeing. I can still do imagery, I can still dream visually, but I can’t see. So, likewise, neurons in the primary visual cortex are important in this sense, but they’re not sufficient for consciousness. And consciousness has to arise from discrete coalitions of neurons in a higher part of the brain, in a higher part of cortex.

And this has interesting consequences. You can test this. And so our claim was that the NCC . . . that visual consciousness does not arise, is not . . . that this part of the brain is not sufficient to give rise to visual consciousness, that visual consciousness has to be generated in higher parts of the brain. And there’s lots of evidence for that, in particular, the most striking is evidence from monkey electrophysiology, from recordings done by Nikos Logothetis and his colleagues, where you can see that the monkey responds. We do experiments where you manipulate, just as I mentioned to you, where you manipulate the relationship between what’s on the retina of the monkey and what the monkey sees. The monkey doesn’t see a stimulus, although physically it’s still present on his retina, and you see literally millions of neurons are firing to this unconscious stimulus, it’s a stimulus that the animal doesn’t perceive and doesn’t respond to because it’s perceptually suppressed, yet there are millions of neurons in primary visual cortex that still respond to this unperceived stimulus.

And there’s also some nice evidence from human . . . there’s also now some recent evidence that other primary areas, like the primary auditory cortex or the primary somatosensory cortex—all you can have a lot of activity in those regions without being at all conscious about them. So it may be possible that none of the primary sensory cortical regions actually are sufficient for consciousness. You could say, Well if that’s true, you know, that’s sort of a minor detail—who cares? Well that’s, I think, very interesting because it suggests that not any neural activity is sufficient for consciousness, in fact, not even any neural activity in [the] cortex is
sufficient for consciousness, that consciousness is something discrete, you know. There are two sets of ideas as John will tell you more about later. One sort of holds that it's impossible to assign the genesis of something like visual consciousness to any one specific set of neurons, that consciousness is a global, holistic, collective, gestalt-like property of the brain, and it's silly to think that it arises from discrete groupings of cells.

Now our intuition is based on sort of the model of twentieth-century biology, which is all about specificity. And you saw that yesterday, for example, in the talk by Richard Axel where you have amazing amounts of molecular specificity at the level of the antennal lobe in a fly. The same overall story is going to be true for cortex and for the generation of consciousness. There are going to be discrete biological mechanisms in the brain. They will involve discrete groups of subtypes of cells that talk to each other in particular ways, and they will give rise to specific types of conscious sensation. While other neural activity does not give rise to conscious sensation, and this is the activity involved with all these Zombie Systems, all these things when you're driving home, you're lost in thought, suddenly you realize you . . . sort of . . . you wake up and you're in your garage while you were thinking about your latest paper that just got rejected. And so here you had to do very complicated activity, we know that there's complicated activity, you had to stop, you know, you had to look at traffic lights, etcetera. We know this involves cortex, so again we know that it can't just be any cortical activity, it has to be specific types of cortical activity, you know, in a particular mode maybe, in a particular region of the brain that gives rise to conscious sensation.

Explicit Representation

The other thing that we think is absolutely essential for the NCCs, an explicit representation underlying every discrete conscious percept, like conscious percept of my voice or of these colors, it's got to be an explicit representation at the neuron level. But what I mean by explicit—this is a picture I've borrowed with Bill's permission from one of his earlier papers—so this shows you an explicit representation for depth and for motion in an area called MT or V5. This is a part of a cortex of a monkey where you have a whole—this is part of cortical layer here, so this is roughly two millimeters, and here you have an entire set of neurons from layer II to layer VI that all sort of more or less code for motion in this direction. And over here neurons code for motion in this direction or this direction or this direction. And they encode this direction of motion explicitly, in other words it's very easy for you as a postsynaptic neuron, as a neuron network or receptor, it's very easy for you to read out that information explicitly. Likewise, if you go in a higher part of the brain called inferotemporal cortex, where we know from monkey recordings a lot of neurons that encode faces, that encode faces in a very explicit manner. In other words, for a postsynaptic observer, a postsynaptic cell, it's very easy to decide based just looking on the output of these neurons whether or not a face is present. You take the counterpart to explicit, it's an implicit representation. So everything I know about the visual world is already present in the retina, because that's the only
source of information. In fact, it's already present at the level of photoreceptor. But things are not made explicit. At the level of the voltage in the hundred million photoreceptors in my eye, this information is not made explicit. Information for faces is only made explicit at a high-level stage. And the idea is that the key apparatus . . . is that the neural correlates of consciousness has to be based on such an explicit representation. Why? Because we're directly conscious of something, we don't have to—one of the remarkable features of consciousness is that I'm directly conscious of it. That's what it means to be conscious of something. And so therefore there has to be a direct neuronal counterpart, there has to be an isomorphism, in other words, between anything in my conscious representation, between any attribute of qualia and the underlying neuronal representation.

Then there's the ideas that our . . . we have to ask, How do we experience the world as evolving in time? Are experiences evolving continuously or do we experience this evolving discretely? Certainly if we introspect we experience a smoothly changing world. There is evidence that that's an illusion, that similarly, like the way you experience motion in a movie, which we know actually is incorrect, there's nothing that moves actually in a movie, right? In a movie what you have, you have a discrete image here, and then at 72-Hertz frame rate you have a discrete image here, here, and here, and if you do it quick enough that creates the illusion of motion. Likewise, it is possible that perception in the brain also occurs in these discrete snapshots. These snapshots can be a variable period between, let's say, 50 milliseconds and 120, 150 milliseconds, depending on all sorts of circumstances, the saliency of the stimulus, etcetera, that perception actually occurs in these discrete episodes.

We've written about this last year, and then we were struck by something that Oliver Sachs communicated to us, and he has subsequently written about in the *New Yorker*, what he calls cinematographic vision, which is under conditions of migraines. And he himself has experienced this, I think I have a slide of that—no. So this is sort of the metaphor, that the way you experience things is actually not continuous, but it's discrete. But if it's quick enough, just like in a movie, you can't tell the difference unless you have an explicit mechanism in your head that tells you, and that actively, explicitly signals this difference. So what happens in cinematographic vision? These are people who have visual migraines, and what they experience is—and here's this very vivid description that he himself has had—that the world is fragmented in time, that you see things. Like, for example, he describes when Oliver Sachs had this migraine, that he sees a nurse approaching him in the hospital, but sort of she—it's like the movie's run too slow, and that's in fact the description that most of the patients use when they have these type of cinematographic visual migraines, it's like seeing a movie run very, very slowly.

Several speakers yesterday and today mentioned attention. What is the relationship between attention and consciousness? Now different from William James, we think that attention and consciousness are actually distinct states, that attention is a general set of mechanisms, there's both bottom-up attention that's
inherent in the input, and there’s top-down attention when I can sort of direct my attention in a trained situation—like in Bill's monkey—I can direct my attention to one or to the other stimulus, even though they are of equal visual salience. Attention is a set of selection mechanisms that enables me to take all the stimuli that compete for my consciousness, which is much more than I can process in real time, and to select a very small subset of those stimuli. And a subset of those stimuli, those are the ones that I'm then conscious of. And we know from a hundred years of visual psychology that there are many, many things outside there—we also know this from our personal experience—that I'm not conscious of at any given point in time. So attention is a set of neural mechanisms that selects, out of all this competing stimuli, you know, that are on my retina, that are on my body, that are in my cochlea, that are sort of in my internal imagery and representations, selects a subset of those and then a subset of those again are the ones that are consciously accessible. If they're consciously accessible then you can talk about them, you can access working memory. And that selective attention is necessary to what Richard Axel yesterday referred to as the binding problem. That if I want to recognize certain objects, particular objects I haven't seen before or if I want to combine property attributes of different objects at the same time, that's when I need selective attention, that's one of its key functions.

**Styles of Neuronal Firing**

Much has been made, including by ourselves, about different types of neuronal firing, so you may know electrophysiologists have characterized under different conditions if you listen to the way neurons talk to each other they seem to have different modes. And sometimes you hear neurons that just fire randomly, rat-tat-tat-tat-tat, like a Poisson; in fact, a Poisson statistic is a reasonably good approximation of that. But then under other conditions you can also hear that neurons sort of fire rhythmically, in various frequency bins, particularly, there's this one called **40-Hertz firing** where neurons seems to fire with some sort of periodicity in the 20- to 30-millisecond range, roughly in the 40-Hertz range. And then furthermore what you can see under certain conditions, neurons from two separate neurons don't just fire independently of each other but seem to appear to fire in synchrony. This is called synchronous firing. And many people have argued that one of the key properties underlying any neuronal representation of consciousness is the fact that neurons that code for conscious—or that express or that are jointly sufficient for conscious—percept are the ones that fire in synchrony. The evidence for that based on monkeys is rather inconclusive. There's some evidence, the evidence is not very good. The best evidence indeed for the importance of synchronized firing comes from the olfactory system in insects, the evidence of Gilles Laurent, so it is more likely that synchronization and synchronized firing is important for certain types of conscious stimuli, in particular when they are competing into each other. In fact, there's now some evidence from Bob Desimone that 40-Hertz synchronized oscillation may be important for biasing correlation. So you have two stimuli—both compete for attention—you can only be conscious of one of them, so you somehow bias and you tend to one stimulus, not
the other, and that one of the neuronal signatures of that bias is actually synchronized oscillatory firing. But that this type of firing therefore often occurs with consciousness, but that it may not be absolutely necessary. In particular when the conditions like you often have in a lab where you just put a single stimulus out on a monitor and the subject is looking at a single stimulus when there isn't a lot of . . . when you don't really need too much attentional-selection bias because there is really no other competing stimulus. So in other words, under certain conditions, there may be this relationship between synchronous firing and consciousness, but it probably does not hold in general.

Types of Behavioral Assays

So what is the experimental program we advocate? Well first of all it has to be emphasized again and again that we think that consciousness is an empirical problem, maybe not a purely empirical problem, but consciousness ultimately is an empirical problem that's amenable to sort of the normal scientific method that we've used so successfully over the past few hundred years. What you need is a series of assays in humans, and particularly in animals, you need a series of behavioral assays where you can be sure that the subject is conscious. Now for us that's sort of trivial, I mean I'm certainly conscious because I can feel that, and assume that most of you, if not all of you, are conscious. I just mean those of you who have fallen asleep, not that you're zombies. But in an animal we . . . I mean most neuroscientists would assume because of the great behavioral similarity between sort of a typical subject, if you take a typical undergraduate subject and experiment and you compare that against training an animal in a similar task then you'll find a great deal of—particularly if these tasks do not involve high-level, you know, knowledge, obviously animals don't talk, but involve simple things like vision—you find a great deal of continuity between the behavior of animals—in particular nonhuman primates like macaque monkeys—and humans, the brains is very similar. If I give you a little cubic millimeter of a macaque brain, a little cubic millimeter of a human brain it's very, very difficult to tell them apart. We've got much more brain, but the basic structure, the basic neural cell types, etcetera, are very similar. So therefore, by and large, neurobiologists sort of assume that animals share at least some aspects of consciousness with us, probably not self-consciousness. Self-consciousness seems to be something that's probably highly elaborate in us that maybe only chimps or some, you know, the great apes share with us, but certainly the way most animals, certainly I would believe that most mammals see the world, hear the world, smell the world is probably very similar to us, up to and including the conscious perception of these stimuli. So what you would like, you'd like to have a set of behavioral assays that you can use in monkeys or even other animals that are more amenable to genetic interventionist protocols, such as a mouse, or maybe even if you want to go lower, maybe even something like a fly. And that ultimately — and you need that because doing fMRI and doing single cell, doing psychophysics in humans is great, but ultimately you need to intervene with the system, you need to perturb the system selectively. Cortex is by far the most complicated system for its size in the known universe,
and the only way we're going to unravel the circuit is by sort of taking, you know, going inside the brain, taking the circuit apart and then putting it together again to understand the genesis, origin, and function of consciousness. And obviously we can do that in animals.

One set of assays seem to involve a whole set of different—in humans and in patients—tests that involve keeping information online for at least a few seconds. There's really no convincing evidence that you can do that sort of task, you know, I give you numbers or, you know, a patient has to remember the orientation of a stimulus or the color of a stimulus, there's no evidence that you can do that, or patients can do that, you know, if you have to keep this information online for a few seconds without having access to consciousness. So probably a good sort of set of, you know, like a [inaudible] test, a good set of behavioral assays would involve tests that involve the subject having to keep information online for at least a few seconds.

So we are pursuing a number of paradigms, mainly psychophysics and fMRI, but yesterday night I decided just to focus on two, given the brevity of time. So one involves mice and the other one involves humans.

Recording Single Neurons in Conscious Humans

So almost everything we know about electrophysiology we know from electrophysiology of animals, for obvious reasons. Now there are, occasionally, rare occasions when you can record the response of individual neurons in humans—that's during surgery. So we work with a neuroscientist and a neurosurgeon, Itzhak Fried at UCLA, and the background is that these are epileptic patients, these are patients that are resistant to conventional pharmacological treatment, the drugs don't work anymore on them. And then what neurosurgeons do, it's done throughout the world, it's a quite successful operation, then you go in and surgically remove the part of the brain that gives rise to the foci, which more often than not tends to be in the medial temporal lobe. You remove, for example, a piece of hippocampus. And then, you know, the incidence of seizures is dramatically lowered. In some patients the incidence goes to zero, in most patients it's dramatically reduced. Now in a subset of these patients, roughly thirty patients, on the order of one patient every month, from the outside looking at the etiology, looking at the behavior, looking at the EEG, or even looking at the structural MRI is not sufficient to tell you where is, actually, the foci in the brain. So then what you do, you put in electrodes, you put in up to 12, ten to 12 of these microelectrodes, so these are big electrodes, these are 1.2-millimeters thick, these are polyurethane, very flexible polyurethane, and then you have these platinum-iridium leads. So here you can see one inserted into the hippocampus of this patient. And then the cortex is sealed up, the burr hole is sealed up, and then the patient has, like I say, eight to 12 of these electrodes in his or her head, and then is monitored on the ward for let's say three to five to seven days, it's monitored 24/7. And then when the patient has seizures, in fact the patient's released after he has three to
five seizures, that's typically the number that's sufficient to enable the neurosurgeon to then triangulate where the foci is. The electrodes get taken out, and then that foci is taken out and that's the end of the operation.

Now, in fact what Itzhak Fried and his collaborators did already a number of years ago, they hollowed out the inside of this polyurethane and now they inserted these metal wires. These are just like Bill showed you; now, conventional microelectrode isolated until the last forty micrometers, and they're very thin platinum-iridium metal wires essentially. So typically we use nine wires here and there are eight to ten of these, so we have on the order, and we have 64 preamplifiers, so we can now record from on the order of thirty to fifty neurons for days at times, for two to four days, in the head of a patient. So there are some advantages and some disadvantages to this scenario: The advantage is with the permission of the patient we can do experiments. We can directly query the person about his or her conscious state, we can directly ask you did you see that, did you not see that? And also the patients can do things without training them for many, many months. In a typical monkey experiment you have to train the animal for a very long time and reward him suitably in order to get the monkey to perform or the animal to perform the task that you want the animal to perform. In humans that's of course much easier. The drawback is it's not nearly as well-controlled as in an animal situation, and there are many things we can't do. We can't, for example, get the person to do the same experiment hundreds of times, just because the patient isn't willing to do that and will literally go asleep.

[A portion of the transcript including unpublished results has been removed at the request of the speaker.]

So then what we can do if we record from any of these neurons, like I said, typically now we get between thirty and fifty of these neurons, although many of those are not visually responsive, we can do what people called decoding. And in this case it's just a very superficial linear discriminant decoding, so, in other words, we can ask I'm now listening to, let's say, seven neurons from this one batch of electrodes, all of which seem to respond to various visual images. So on this one trial where I showed an image for one second, how well can I—using some objective criteria—how well can I decide which of the twenty stimulus was present? So once again I show the patient, let's say, twenty different pictures, and now I want to do decoding, and the decoding question is based on the firing of just this set of neurons during this one trial. How well can I discriminate whether it was the picture of, you know, the picture of a person or the picture of that animal, or whether it was the picture of the spider, or picture of the eagle? So we can get these probabilities. This is in different sessions in these patients. This is from the last patients where we plot some of the data. Let's see, the blue is chance level, so if I'm just guessing this . . . so here's a probability, these are just different trials in different patients, this is a probability from 0, 20, 30, 40, 50, 60, 70 percent. One hundred percent is perfect. Chance here varies because it depends on how many stimuli we show. So for example in this particular case chance is 50 percent, so if
I'm just guessing I would get 50 percent. And so this is what a . . . the light blue is the performance of a neuron if I look at all the spikes over the one second when the image was present and the one second following the image. So I can do much, much, much better than chance. I can do even much better if I just focus on the burst of spikes. Very often we have cells that when they respond, particularly when the person recognizes the object, they seem to fire in bursts, and in a burst of spikes, a quick, you know, five to seven spikes in one hundred or two hundred milliseconds, that burst of spikes always occurs between three hundred and six hundred milliseconds. So we just look at that information, we just look at the action potential between three hundred and six hundred, then we can do very often much, much better than looking at all the information. In other words, the information is not just uniformly spread across the interval, but it seems to be specifically, there seems to be much more information in this one particular burst interval.

[A portion of the transcript including unpublished results has been removed at the request of the speaker.]

Here's a cell that seems to be selective to very different pictures of Clinton, doesn't respond . . . let's see here, here, and here, so again this is the response to stimulus where it came on here, came off here, and the horizontal bar tells you the background rate. So the neuron fires on average maybe two to three times a second, but clearly fires very, very strongly to these three different pictures. It's a remarkable invariance property, because if you look at the level of pixels, you know, the relationship between here, which is a gray-scale image where his face takes up almost the entire image, to here, where it's color and is much smaller, to here, where the face is even smaller, and you know, there are other individuals present, is really . . . so at the pixel level these images are very different, but the neuron seems to respond in a very invariant manner to it. We find that quite a bit in these areas. Typically we have out electrodes in the medial temporal lobe, which is amygdala, hippocampus, parahippocampal gyrus, and interrhinal cortex.

**Motion-Induced Blindness**

So this is one particular experiment that we did: What you should do, you should look at the fixation, you should fixate and not try to move your eyes, and just look at the fixation cross in the center. It's a nice illusion discovered by Bonneh four years ago. It works best if you don't move your eyes, if you keep your eyes as still as you can on the cross. There's no way we can make this darker, right? Do some of you see something—yes, no? Oh thank you, excellent. You see things disappearing? Okay, so this is one of these illusions I mentioned, this is called motion-induced blindness, like I was saying, was discovered four years ago by Bonneh et al, and it's one of the many illusions that visual psychologists have now to manipulate what you see, to manipulate the relationship between what's physically present on the screen and what you see inside the privacy of your head. Here the point is, although those yellow squares are present all the time—they're all the time present on the screen—sometimes you see them, sometimes you don't
see them. When you're seeing you're conscious of them, they're yellow, they
remind you, I don't know, of the yellow sun, they remind you of yellow sunflowers
of Van Gogh, I don't know, whatever else you have, what memories you have
associated with yellow. You're conscious of them, you can talk of them to your
neighbor, you can use them to do planning. And when they're gone, they're just
gone, you don't see them anymore, you're not conscious anymore. So the question
is where's the difference in your head? It's a very simple question, right? Where is
the difference in your head? What's the difference in your visual cortex between
when you see them and when you don't see them?

And, for example, you can use these sort of stimuli, this is just one of many such
illusions, like rivalry flash suppression. All are instances of that, where you can
take the footprints of consciousness, right, because now you can ask every time
when you consciously saw the yellow spots, you could do it in fMRI, and has been
done in fMRI, Nancy mentioned that, she has done something like this, when you
physically see the yellow is there a region of the brain that's active in fMRI, or are
there single neurons that fire when you see the yellow? And if you don't see the
yellow but the yellow's still present on your eye, are the neurons, is that neuron,
does the neuron stop firing? If it does then it has its close correlation between
visual consciousness and firing. Of course that's just a correlation, it's not a
causation yet.

For example, you can show, we've shown recently, that you can now—I mean you
all know what a—I assume most of you know what an afterimage is, right? So, for
example, if you stare at the yellow for a long time and then I flash you a black
screen you're gonna see a ghostly blue image superimposed. Why blue? Well
because yellow gives rise to blue afterimage. If you look at a red image for a while
you get a ghostly green afterimage. For example the extent and the duration of
your afterimage doesn't at all depend on whether you consciously saw the
stimulus. So in other words you do not need to see a stimulus in order to get an
after effect. Certainly you don't need to see a stimulus in order to get an afterimage
of the same duration and the same intensity, and you also don't see a stimulus to
get an orientation-dependent after effect. All those things preceded consciousness.
So what that tells us there are discrete stages in the visual system, in the visual
hierarchy, and that at one stage you have the neuronal correlate of these after
effects, for example, the afterimage that's probably in the retina, or orientation-
dependent after effect in V1, and that visual consciousness has to arise at a higher
stage. So again that's important because it tells you it's not just one holistic thing,
but that there are discrete processing stages and consciousness seems to arise at
or beyond a particular processing stage in the brain.
Flash Suppression

So now, in principle, we can do the same in a patient. So this is a similar version, it's just more difficult to demonstrate in an audience, which is why I'm using the other one. Flash suppression is a similar illusion, was discovered and characterized by Jeremy Wolfe at MIT, and essentially involves the following: I show you an image in one eye, so let's say you have an image of my hand in one eye, and then I don't know, and you clearly see that, and then I flash up in your other eye I flash up the image of the watch. Now both the image of my hand, if I do it carefully the image of my hand and the image of my watch are simultaneously present, but perceptually the new input trumps the old input, and what you see is only the watch, you don't see the image of the hand, although physically the image of the hand is still present in my right eye. Okay?

And there were some beautiful experiments done by Nikos Logothetis and David Leopold in a monkey where they showed already in a monkey that there are cells early on, in primary visual cortex, that don't care about—that seem to fire to the stimulus whether or not the animal saw the stimulus, whether or not the animal behaved to the stimulus. So you can get a monkey to train in flash impression just like you can do, you know, its behavior is very similar statistically to the behavior of a human, and you can see in a monkey that cells in V1 fire—fire very vigorously, even though the physical stimulus is present. So again, that supports the idea that primary visual cortex, that's not where visual consciousness—the neurons in primary visual cortex are not sufficient for visual consciousness.

So we can do the same in patients, so this is one neuron. So here we show Curly, we showed Curly because we found neurons that responded to Curly in this gentleman—well, in fact, this is the neuron that responds to Curly, so here we put Curly on, nothing in the left eye, the patient recognized Curly, incidentally, and here you can see—again you see this burst of activity around three hundred milliseconds, and the person sees Curly. Now you flash on a grating in the left eye. Perceptually this grating suppresses the image of Curly and what you see—you don't see a superimposition, you actually only see, it's quite striking, you only see grating. It's very reliable, easy to set up in the lab. And after a suitable delay, this response here [is] statistically no more different, statistically not different from the response here. So in other words, the neuron behaved statistically, at least to the best of our statistical abilities. The firing here is no different than the firing here.

Here we do the opposite case, we show the grating, the neuron doesn't fire to the grating. This neuron really tends to like only Curly, this is the only stimulus we found that this one cell happens to like. The person sees Curly, you flash up Curly in his right eye, perceptually this image suppresses that and you see Curly. Now notice here, this input and this input are the same physical input, the only difference is here the person says—I mean I know it, we asked him—sees a grating, and here, he sees Curly. So again, here the neuron fires very strongly. So
in this case the neuron correlates with the visual consciousness of the person. So we can do that in category-specific cells, we can do that in cells—here we averaged a lot of cells in a couple of patients that are category selective, here we did it for neurons that are selective only to individual images, so I can just show you the average response. So when I put on the preferred stimulus of that neuron, like Curly or Clinton, the person saw the stimulus and the neuron fired, then here the preferred stimulus was still present, the stimulus that the cell responds to it was still present but was perceptually suppressed, and the neuron fired only weakly. And here the opposite is the case, that here the stimulus isn't present, the neuron doesn't fire. Here the stimulus is present and is perceived, and the neuron fires very strongly. So perceptually this condition here, and this condition are the same. Here you only have the preferred stimulus present, here you have the preferred stimulus present in one eye that suppresses the other stimulus in the other eye, and you can ask, Is anything different about the neuronal response between here and here? And we did that and the answer is no, neither the duration, nor the amplitude in terms of the number of spikes, nor the peak response is different between those two cases. So again, statistically, we cannot tell apart based on the firing of this neuron; sorry, the neuron cannot tell apart, or at least the spiking of the neuron does not distinguish between the two situations that are phenomenologically the same. That I have only Curly present, or I have Curly and the other stimulus present, but Curly suppresses the other stimulus. In both cases, to the observer, they look the same. I see Curly, and the neuron signals that likewise.

So now this is, of course, just correlation, like a lot of electrophysiology. All I can tell you is that there's a nice correlation between what the human said he saw and the behavior of the neuron. By the way, this was true for two-thirds of the neuron in this part of the brain followed the percept, one-third of the neuron just didn't fire at all or fired much reduced when two stimuli were present. We never saw a scenario where a cell fired to a perceptually-suppressed stimulus. In other words the unconscious, you know, the Freudian unconscious if you want, wherever it is, it's not present in the firing rate of neurons in the medial temporal lobe.

**From Correlation to Causation**

So there's this nice correlation. Ultimately, you want to move to causation. Now that's not impossible in human neurosurgery, and possible scenarios one can imagine among other stimulation, occasionally—or, in fact, quite commonly—neurosurgeons do stimulate parts of the brain during surgery to make sure, to understand where they are and where the language areas are and where the motor regions are. So it's not implausible that you can think about a protocol, there are a number of obvious ethical and practical questions involved, where you can think of a protocol where you can directly stimulate, you know, a bunch of neurons that seem to code for faces or for animals to try to see: Can you actually switch—in a reliable way, can you induce a percept or can you reliably switch the percept of a human in order to begin to take the jump from correlation to causation?
Of course in animals, this is much easier, and so that's where we're exploring a totally different paradigm. So this is work we've been doing with my good colleague David Anderson at Caltech, and Michael Fanselow at UCLA. And David and I have two postdocs, C. J. Han and Colin O'Tuathaigh, who actually did all the work. So this is based on a paradigm by Clark and Squire. So there are many forms of associative conditioning, some—have been, since a long time—require conscious awareness of the relationship between the CS and the US, other ones do not. In this Science paper that they wrote five years ago, they looked at—in humans—they looked at eye-blink conditioning, and they argued that this form of conditioning where you have a tone, beep, and at the end of the tone you get a puff of air to your eyes. Now that's very annoying and you blink. If you do this a hundred times and you just get the tone you immediately start blinking, your deep cerebellar nucleus, your brain has learned to anticipate, when this tone comes I've been conditioned to expect a puff of air to my eye and I blink. This is a more complicated form of conditioning, it's called trace conditioning. Now you have a tone, beep, you can also make it long, they've also tested that, you have a tone and then you have an intervening trace period of one second, and only then does the puff of air come. Now this is, as I mentioned to you before, you know, as an assay to involve consciousness, this is one of them, I think, because it involves this intervening period. And so now the animal, the human, there isn't just the tone and the puff of air but there is an intervening period and you have to keep it dynamically online. And they showed some nice evidence in humans that this requires awareness, this requires, here people need to know that there were tones and there were puffs, and the tones always preceded the puffs in order to express conditioning. Here, whether or not people were aware of this made no difference; here, the subjects were always conditioned. We repeated the same thing in humans using shocks, we did it using electroshocks because we'd like to move to mice, or we're moving to mice, I'll show you that now, and in mice you can do field conditioning, this is called field conditioning, much more reliable, it takes many fewer shocks than puffs of air.

So what we did is the following: So this is now in mice. So we have delay conditioning, here we have a tone, beep, for 16 seconds, I'll show you a movie, and then the floor of the cage is electrified. That's delay conditioning. And here's trace conditioning, here there's an eighteen second trace period between the end of the tone and the shock. And so the animal has to keep this dynamically online. Then we try to distract mice. One of the ways Larry Squire showed that you require attention and awareness in humans—he distracted humans. So we do the same and we try different things, and here we distract them by flashes of light, a bit similar to what Eric Kandel was telling you yesterday about in his case. So here there are flashes of light, and then we do the same thing, we do trace, delay conditioning and trace conditioning. So these mice have never been shocked. This is the very first time we shock them.

Okay, so it's working now. So there's this tone, which you didn't hear, because I didn't put on the audio, and you'll see what happens. There's these flashes of light,
and soon you'll see what happens to these critters. Okay, so now for two seconds
the floor was briefly electrified. And now, you know, they're all very nervous. And
we do this six times. And this was the very first time. We do this in day one. Then
on day two we take them into different context, there's no context-dependent
conditioning, and then we test them. So here you have two sets of mice from one
conditioning and two sets of mice from a different paradigm. One set was
distracted, the other one was not distracted, you'll see which one. And the
measure of conditioning we use is freezing, how much do the mice freeze, and
you'll see—okay, there's this tone—okay, there's no tone here. Okay, so I can just
tell by the light, the tone is on and you can see these mice totally stopped moving,
the mice are somewhat reduced moving but they still move, so this is called
freezing. And you can measure the amount of freezing by doing behavior using a
videotape and measuring every second this mouse is frozen, that mouse just
stopped freezing, here the mouse stopped freezing but then it goes back to
freezing, these mice freeze much less. The difference is these mice on previous
days were exposed to the trace-delay paradigm with distracter, these were
exposed to the same shocks, the same tone, but here they were distracted. So
here you can see the behavior average, so this is delay conditioning with and
without distracters, statistically there's no difference. Here's a very significant
difference between the mice that trace conditioned that were . . . so these were not
distracted and these were distracted.

So the other way, as I said, we now want to begin to move to interventionist
protocols, so what we then did—and I guess I'm running out of time—we did
pharmacological lesion to remove the anterior cingulate, which we know from our
functional imaging experiment is a part of the brain in humans that's involved
specifically in trace conditioning, but doesn't seem to be so much involved in delay
conditioning. And if you do that, if you remove the ACC in these animals, if you
remove this in animals then you get this very nice behavior. So these are the
normal animals or with sham surgery or V1 surgery, these are the animals with
ACC lesion. Makes no difference whatsoever in delay conditioning, but essentially
eliminates all of trace conditioning as compared to shock only. And it doesn't
interfere with context conditioning.

Conclusions and a Warning

So, to finish, what we could show here in the mice—we have a nice model, similar
to what Eric Kandel has, we have a nice model for attention, possibly if you believe
the link to humans with awareness—in a sense that we can show two forms of
conditioning: trace and delay conditioning. If we distract the animal it seems to
interfere specifically with trace but not with delay conditioning. And if we remove a
part of the brain, ACC, that in humans in involved in something similar, you can
again eliminate trace conditioning without interfering with delay conditioning or
context-dependent conditioning.
Let me give you one last slide. So people say well, this is all very fine, nice and fine and you know, this program of finding the neural correlate of consciousness is interesting, but you know, will that explain it? So let's say it's layer V cells in inferotemporal cortex that project to prefrontal cortex and back that are sufficient for consciousness. How does that explain something? And there's this wonderful quote I found, and this is Bateson, who was a very famous English geneticist, and he reviewed a book of Thomas Hunt Morgan, who was at the time, here, I believe, at Columbia, during the war, and in this book he wrote about his evidence based on flies, that genetic information was stored along one-dimensional strings. And so this is what he wrote, "The properties of living things are in some way attached to a material basis, perhaps in some special degree to nuclear chromatin." We know that's true now. "Yet it is inconceivable that particles of chromatin or of any other substance, however complex, can possess those powers which must be assigned to our factors or gens [sic]." That's his spelling. "The supposition that particles of chromatin, indistinguishable," that's of course incorrect, "from each other and indeed almost homogeneous under any known test can by their material nature confer all the properties of life surpasses the range of even the most convinced materialism." The trouble here was that they thought they understood chemistry. And so by their test at the time in the early 1900s they couldn't distinguish different strings of one-dimensional information, in fact they couldn't even imagine, they didn't even have the concept of sort of specific-marker molecules. Hemoglobin wasn't characterized until later. And so they couldn't imagine the amazing complexity, the amazing amount of information you can specify in one-dimensional strings of nucleotides.

Same thing, we're only beginning to explore cortex, cortex is amazingly complex and we really have very little understanding yet how complex it is. And so, I think, one should be very careful. Many people think, you know, the study of consciousness clearly can't be addressed by scientific methods and, you know, it requires extra scientific things, and I think we should just be very careful of asserting that, given that we've made this mistake several times before in our own intellectual history.

Thank you very much.

**David Cohen:** I'm afraid we don't have time for questions. Perhaps you can talk to Dr. Koch after the session is over.
Brain and Mind
May 14, 2004
Session III: Biology of Mind

John R. Searle, PhD
Consciousness, Causation, and Reduction

Introduction by David Cohen

David Cohen: Our final speaker of this session is Dr. John Searle, who's Mills Professor of Philosophy of Mind and Language at the University of California, Berkeley. Dr. Searle was educated at the University of Wisconsin and Oxford University, where he was a Rhodes scholar. All of his degrees are from Oxford, and that's where he subsequently began his teaching career, but since 1959 he has been on the faculty at Berkeley.

John is widely published, including 15 books, and his work has been translated into 21 languages. Among his well-known books are Speech Acts, Minds, Brains and Science, The Mystery of Consciousness, Consciousness and Language, and there are many more. He is recognized internationally as one of the most distinguished contemporary philosophers of mind and language. Indeed, his contributions have been recognized by numerous prestigious visiting lectureships around the world and by several honorary degrees. It is indeed an honor to welcome John Searle, who will speak today on consciousness, causation, and reduction. John?

A Dualistic Conception of Consciousness

John Searle: Thanks a lot. It's a great honor to be here and yesterday when we were being graciously welcomed, all us foreign visitors, it suddenly occurred to me I was probably a Columbian before anybody else here. I was a student at an experimental school run by Columbia University in 1945. Now those of you who antedate me, I'm glad if there are some of you, but for the rest of you I want to welcome you all to Columbia. It's a great place. That is, I have to say that in the ninth grade at Horace Mann-Lincoln School was the most intense intellectual environment I have ever lived in in my life. And I'm continuously trying to recreate it in various universities.

Anyway today I'm going to talk about . . . well, I might as well tell you the awful truth, there is a single overriding question in contemporary intellectual life. It is such a vast question that I think unconsciously most professors do everything they can to prevent their students from finding out that they don't know the answer to it,
and in fact they'd rather not even think about the question. So let me tell you what
the question is. We have a pretty good idea now about how the world works. We
don't know as much as we like to pretend we do, but we know quite a lot about
atomic physics and chemistry and molecular biology and even evolutionary
biology. So we have a picture of the world, [and] it's basically made up of physical
particles in fields of force. Now, of course, particles isn't right, that's the wrong
word for points of mass energy, but nobody's listening so let's just say particles,
what the hell. The world is made of physical particles in fields of force. These are
organized into systems. Some of them have got big carbon-based molecules with
lots of nitrogen, hydrogen, and oxygen. And over a period of about three to five
billion years they evolved into us.

Okay, now that's the world, that's how the world really is. But now we have a
certain conception of the world and our place in it. We are conscious, we have
intentionality, we have free will, we have rationality, we have language, we have
society, we have ethics, we have a self-conception which we're pretty reluctant to
give up on, we're pretty fond of that self-conception. Now here's the question: How
do we reconcile what we know about the world—so to speak, the basic facts—with
that self-conception? And that basically is the dominant question in philosophy,
though not all my colleagues agree with me about that, and I think it's the dominant
question in a whole lot of other fields.

Now the reason that I'm here today is that some parts of that question,
unfortunately not all of it, but some parts of that question, I think are going to get a
scientific answer. And one absolutely crucial part of that question is
consciousness. What the hell is consciousness, and how does it fit in with the rest
of the world?

Now part of the problem we have here is we can't approach this innocently. We're
stuck with a tradition that goes back over 2,000 years of thinking, well we really live
in two different worlds, we live in the mental world of our inner soul, and we live in
the physical world of physical particles. And so we start off with this dualistic
conception. We're brought up on it, it's enshrined in our popular culture, we sing
songs about our body and our soul, and we even have sayings about how the body
is willing and the flesh is weak, or maybe it's the other . . . no, the mind is willing,
the flesh is weak, I always get mixed up. But anyway one of them is willing and the
other one is weak. I feel pretty weak on both halves of the story. But in any case
it's hard to shake dualism. And I've listened to some of the most heavy-duty brain
scientists in the world over the past couple of days, and you'd be surprised how
often they resort to a dualistic vocabulary, how often the little homunculus rears its
head, as I decide to do such and such, and I shift my attention. And I don't know
that we've got another way to talk about it.
Defining Consciousness

Okay, now you might think, all right, so consciousness is a scientific problem like any other, so why don't we just let these guys get on with the job, give them their research grants, give them their graduate students and turn them loose. And basically that's what I'd like to do, but unfortunately they've all had a philosophical upbringing, too, even though a lot of them don't know it, and they are prone to the same kind of mistakes that the rest of us make. So I'm going to try to correct a few of those mistakes. I'll tell you, nothing is more unwelcome in any audience whatever than to say cheerfully, "I've come here to correct your errors." But anyway this is what we're going to do.

First of all, one of the things everybody always says is, "Consciousness is very hard to define, we can't define it." Actually I think it's rather easy, and I'm going to do it in a couple of minutes. You see, you've got to distinguish between the scientific definition that comes at the end of the investigation where we now know how it works, and you've got to distinguish that from the common-sense definition that you start off with, the aim of which is to identify the target. So the scientific definition of water is H₂O. The common-sense definition is it's [a] clear, colorless, tasteless liquid and falls out of the sky in the form of rain and it flows in streams and rivers. Now with consciousness we're still in the clear, colorless, tasteless liquid stage. But we don't know the scientific definition. That's what we're working toward. But the common-sense definition I think is rather easy, so here goes. Consciousness is defined as those states of sentience or feeling or awareness that begin in the morning when you wake up from a dreamless sleep, and they continue on all day long until you fall asleep again, get hit over the head and knocked unconscious, or go into a coma, or die, or otherwise, as we would say, become unconscious. Now that's the target.

Now notice on this definition a couple of features. Dreams are a form of consciousness, they're sort of touchy feely stuff that goes on while you're sound asleep, but notice also that I have defined consciousness in a way that leaves it open whether or not you're going to have consciousness without self-consciousness. I think you can. I think self-consciousness is a fairly advanced capacity. But a lot of philosophers will tell you no, you can't have consciousness without self-consciousness. So I'm leaving that issue open. But it's these feelings of sentience or awareness that are the target of our investigation.

One of the things I want to call your attention to is these things never just come to us in an isolated form, but we always have a particular feeling within a whole conscious field. So right now, I don't just hear the sound of my voice and taste the water going down my throat and see a sort of blur because the lights are blinding me and feel the pressure of the shirt on my neck, but I have all of that in a single unified conscious field, and I think we ought to take that seriously in our investigation. And by the way, one of the best ways to study any of this stuff is in
the pathological cases, and this is why the split brain patients are so interesting to us. You remember these sad people. They have their corpus callosum cut, and the result is that sometimes they behave in a way as if they had two independent centers of consciousness, as if there are two guys living inside this skull. But in any case in normal, nonpathological consciousness of the kind that you and I are having right now, you don't just have conscious experiences, you don't just have qualia, but you have your qualia as part of a unified conscious field.

All right, so that's our definition of consciousness. Now what's our philosophical and scientific problem? Well, to put it in very simple terms, we want to know how the brain does it, how exactly does the brain cause consciousness, and how is it realized in the brain? Now we have an obstacle to that because, as I said, we have this tradition that says consciousness isn't really part of the material world, and we have a vocabulary that distinguishes between materialism, on the one hand, that says there isn't anything in the universe that isn't material, and dualism, on the other hand, that says no, we really live in two worlds, the mental world and the physical world.

Now I think both of those views are false. I think materialism is wrong in saying that there isn't anything in the universe except, so to speak, an objective, third-person ontology—ontology is a fancy word, it means what exists, mode of existence—there isn't anything in the universe that doesn't have a third-person or objective ontology. And the dualists insist that no, there's got to be—they love this phrase—over and above the material world. Now I think they're both wrong, but they're both trying to say something right. The materialists are trying to say, as I said a few moments ago, the world consists entirely in physical particles in fields of force, and the dualists are right in saying all the same we really are conscious, we really do have conscious states and feelings and you can't get rid of them, you can't reduce them to something else.

Okay, so what I'm really going to try and do in this talk is show you how you can reconcile those two if you avoid certain traditional mistakes. Now I'm going to do something I've never done before, and that is I'm going to use this dreadful apparatus. When I was told they would have this apparatus, I insisted that my lecture notes would be reproduced for everybody because the problem with this apparatus is you can't stuff it in your pocket and take it home, and my attention always wanders when these damn things are on the screen. You know, so many diagrams—I'm busy on the left-hand corner and the guy's already onto the next one. So anyway, you've got something you can take home in your pocket. At least you've got my lecture notes. All right.

**Known Facts**

Now in any investigation I like to start with what I know for a fact, I mean what do we really know? And so I write down what it seems to me we know. Now maybe we're going to have to give up on it, you know, presumably there was a time when
people have written down what we know, the Earth is flat. Okay, so we can't take it for granted that we really know it, but at least these are the data that we would like to explain. Now the first is consciousness is a real feature of the real world. You can't just show that it's an illusion, or it can be reduced to something else, or we can get rid of it. Now I hope that sounds kind of innocent and obvious to you, but I can tell you I've been struggling for about thirty years to convince a lot of my colleagues that that's so. A lot of people want to say, "No, no, no, it's really just a computer program running in the brain, that's all there is to it, or it's really just dispositions to behavior, or really there's nothing going on in there except neuron firings." Now you can get exhausted fighting these guys, but I have to tell you progress does come in the end. I haven't heard anybody get up in these meetings and say, "There's nothing going on in there except the computer program running," so maybe we've got some progress. Anyway, I think this is point number one.

Now the second thing is—I think one of the decisive facts that we know as a result of the past hundred years or so—and that is all of your conscious thoughts and feelings are caused by, and I want to emphasize that caused by, they're caused by lower-level neurobiological processes in the brain. And if you look at the standard textbooks, their favorite level of explanations is neurons and synapses. Maybe that's the right functional level, maybe not. I hope they're right because there's an awful lot of research bet on that that neuron is the right level. But anyway, whatever it is, whether it's the neuron and the synapse or the microtubule or whole maps of neurons or maybe clouds of neurons, we know that the work being done in the plumbing produces all of your conscious experiences. Now that is a stunning fact. It doesn't just mean that the taste of the water or the color red or the blue of the sky, but everything, falling in love or appreciating Beethoven's Ninth Symphony or pick your favorite, feeling the angst of postindustrial man under late capitalism, whatever, whatever is your favorite feeling, remember that is all caused by a whole lot of squishy little things blasting away somewhere inside your skull.

Okay, but now the next question is, Well how do these things exist? And there I want to say I think this is also an obvious fact, though again we have a 2,000-year tradition that urges us to resist it, and that is it's all stuff going on in the brain, that is we ought to think of consciousness as a set of higher-level processes going on in the brain in the same way that we think of the liquid state of the water as a higher-level feature of the system of water molecules. Notice, it's a feature of the system and not of any particular element, so I can't reach in here and say, "I'll find a wet one for you, wet molecule," and similarly I can't reach in here and say, "I'll find a neuron that's thinking about your grandmother." I mean it's not at the level of the individual neuron and synapses, but this is common in nature, that you find a feature of a higher-level system, which is causally explained by the behavior of the elements of the system, even though the system is composed of those elements. There's nothing else in there but the elements; that is, it's just water molecules and it's just neurons in here, but the neurons have this higher-level or system feature.
Now there's another thing that a lot of us are inclined to resist, but I think we have to accept, and that is consciousness is not only real but it actually works. It's a real functioning part of the real world. There's always somebody that will tell you that consciousness cannot really affect the world because it hasn't got any physical weight to it, it hasn't got any electrical charge, it hasn't got any force or mass that can actually exert any pressure. If you want to know why your body moves, you've got to look at the ion channels and the secretions of the acetylcholine at the axon end plates of the motor neurons. And I wish I could say, "Next slide," and show you all that stuff. But whenever somebody tells you that consciousness cannot affect the physical world, my inclination is to say, "You think it can't, just watch." I decide to move my arm and the damn thing goes up, and that we ought to . . . I mean in philosophy, and I think in the sciences, you've got to be astounded by what everybody else thinks is absolutely obvious, but I think we ought to take this seriously, that my consciousness can affect my physical body, and it does anytime I want it to happen. Now notice also there's nothing miraculous. Notice I don't say, "Well that's the thing about the old arm, some days she goes up and some days she doesn't go up." No, it's entirely up to me.

**Approaches to Scientific Research**

Okay. So it seems to me these are the data that we are trying to explain. All right, now let's suppose that we accept those as the data that we're trying to explain. Then it seems to me well why don't we just get busy now and solve it. And the way to do it is the way you solve any of these problems in the sciences: first you find a correlation. And Christof I think is just terrific. And by the way, Christof is too modest to tell you he's just written a terrific book on exactly this subject. So since he wouldn't put in a commercial for it, I'll put in a commercial for Christof's book *The Quest for Consciousness*. If you were my students it's required reading, I would say. So that's the first step, you try to find the NCC, you try to find the neuronal correlate, and that's step one. Step two, you try to find out whether or not the NCC is actually causal, and you do that by the usual methods. That is to say, you find out in an otherwise unconscious subject, can you create consciousness by producing the NCC? And in a conscious subject, can you produce a cessation of consciousness by subtracting the NCC? NCC, remember, means neuronal correlate of consciousness. And then third, you would like a theory, you'd like a theory as to why this particular neuronal structure, this particular NCC, causes this particular effect. For a long time, people thought, "Well, we can't explain why red looks red or why warm feels warm." And I was even told when I first got interested in this a couple of decades ago by famous neuroscientists that science would never be able to do that. It seems to me that's what I'm paying you guys to do. I want to know exactly why red looks red and why warm feels warm; otherwise, you're not earning your paycheck.

All right. So now those are our three stages. You've got to have the neuronal correlate, you've got to make sure that it's causal, and then we want a theory, we want a theoretical account of why it works that way. And if you look at the history of
science that's typically how it works. The germ theory of disease from the time of Ignatz Semmelweis went through exactly those stages, and I see the explanation of consciousness as going through those three stages.

Well, what's our problem? Why do we have so much difficulty? Well, I think in fact that much of the research is based on an interesting mistake. Now you get in a lot of trouble if you're a philosopher and you tell these scientists they're making a mistake, but I'm going to get in trouble anyway, so let me just go ahead and say what I think is the mistake. The temptation is to think if we can find a particular percept, say the experience of the color red, and we find the NCC for that perception, and then we find out how that works we will have cracked the problem, because presumably what works for the color red will also work for the taste of water or the smell of the rose of the sound of music. Now I call that approach the building-block approach, and the idea is you should think of the conscious field as made up of all these building blocks, all of these different perceptions that you have at a particular time. And as Christof pointed out, there's a lot of beautiful research that's done on the building-block approach, and he mentioned Logothetis. And let me just remind you, he gave the example of cross lines and a curly-haired guy. But I remember in one article I read, you show a grid consisting of horizontal lines to one eye and vertical lines to the other eye. And the interesting thing is that the subject does not typically see a grid as a result of that, the subject switches, you have what's called binocular rivalry. Sometimes the subject sees horizontal lines, and other times the subject sees the vertical lines. And now that looks like it's a perfect case, because if you could track the neuronal pathway—which remember is absolutely constant, it's constant throughout—you track the neuronal pathway over the LGN back to V1 and then you track it through the visual system. If you could find the point where it branches—that is, find the point where, although the stimulus is conscious, you're now getting an experience of horizontal lines and then it flips, you get an experience of vertical lines—there must be corresponding change in the brain. That is the model of the building-block research, and as I said there's a lot of beautiful research, not just on binocular rivalry but on gestalt switching and blind sight and a whole lot of other stuff that I really don't have time, and fortunately don't have any slides to show you about all that. But that's the basic idea. If you can crack it for one type of perception, you're going to have solved the problem of consciousness, or at least you're going to solve the major part of it.

I'm pessimistic about that line of research, and let me tell you why. There's one commonality to all this stuff, and that is they always work on subjects that are already conscious. You see, it's only the guy who's already conscious who has the difference between the vertical lines and the horizontal lines. And what I want to know is, How did he get to be conscious in the first place? That is, suppose you adopt a different research strategy and you ask yourself, "How does the brain create the unified conscious field in the first place?" Now think of it this way, imagine you wake up in a dark room where it's absolutely silent. Now I'm struck by the fact you can be totally alert, you can be 100 percent conscious, with absolutely
minimal perceptual input. You have near zero sensory input, you don't see anything, you don't hear anything, maybe you can feel the bedcovers and the weight of your body against the bed, but you can have a completely alert unified conscious field without perception. Then you get up and turn on the light and walk around and brush your teeth. Are you creating consciousness? Well in one sense you are because of course you now have experiences you didn't have before. But I want to suggest here's another way to think of it. We should think of perception not as creating consciousness but as modifying a preexisting conscious field. Think of the conscious field that these guys have and then think of the perceptual inputs as creating the NCC, not for consciousness as such but for that particular percept, that particular modification of the conscious field.

Now there's some research. I call that model the unified-field model because we start with the idea that the unified field is the basic target of investigation. Now it seems to me there are these two approaches. Most of the work being done today is done on the building-block model because it's easier. I mean the techniques we have of single-cell recordings and fMRI seem to work better for particular building blocks than they do for trying to figure out a whole unified field of consciousness, and maybe they will succeed. I mean this is going to be settled by science and not by philosophical argument. It is, after all, a scientific question. Why am I willing to bet on the unified field model? Well very simply, the building-block model would predict that if you had a subject who was otherwise completely unconscious, and you gave him the NCC for the color red, you just triggered the color red, the guy would suddenly have a flash of red and then lapse back into unconsciousness. I don't think that's how the brain works. I think in order to have the experience of red, you've got to have a whole lot of other things going on.

**Typical Philosophical Confusions**

Okay. Now this is an open question, and I just wanted to tell you that as a sort of a lead-in to what sorts of obstacles we have and what kind of research is going on. But now it seems to me, I promised I would talk about some of the standard mistakes that have blocked our progress in this field, about the typical philosophical confusions that we've inherited from our past. And I'm now going to do that. Now I shudder when I do this because I'll tell you the first rule of pedagogy is never write a falsehood on the blackboard. You will see it in the exams if you write it on the blackboard. So I have tried to cover myself by saying [that] they're supposed to be or it's supposed that they can, it's assumed to be, and all that, and whenever a professor says that, that means it's false. Okay, so I used to think it was just the first one, that we were hung on dualism. And I'm now going to do that. Now I shudder when I do this because I'll tell you the first rule of pedagogy is never write a falsehood on the blackboard. You will see it in the exams if you write it on the blackboard. So I have tried to cover myself by saying [that] they're supposed to be or it's supposed that they can, it's assumed to be, and all that, and whenever a professor says that, that means it's false. Okay, so I used to think it was just the first one, that we were hung on dualism. And you have to remember it's not inevitable. We think dualism is somehow built into the structure of our language. I have a friend from Kenya who says that the mind-body problem can't even be stated in his native language, they don't have words . . . oh gosh, if only we had that language. But in any case, on the other hand, ours is pretty pervasive. I once gave a lecture on this subject in Bombay, and I was on the same platform as the Dalai Lama, a great honor. And I was, well I won't say appalled, but I was
taken aback when he got up and said, "We are both a mind and a body." And I thought, "Well Descartes' influence has either spread or maybe he's tapping into some more universal mistake that we're tempted to make." Anyway, I think this is the biggest mistake and I'm going to say more about it in a few minutes. But I want to say just get out of that hang-up of supposing that there are these mutually exclusive realms; that because consciousness is subjective in the sense that it has this first-person mode of existence, it only exists insofar as it's actually experienced by human or animal subjects; and that therefore it can't be part of the ordinary physical world we all live in. That's the main message I want to get across, is that the ordinary physical world we all live in contains consciousness as one of its features, and if we're embarrassed to say it's a physical feature, well then let's just get rid of that terminology and just say it's a biological feature.

The two names that we're stuck with, materialism and dualism, seem to me inadequate, and so I invented another name under pressure. I was giving a lecture in Reno, Nevada, and some guy raised his hand and said, "What's the name of your theory?" I didn't have a name. But I thought, "Well you can't have a theory if you don't have a name," so I said, "Mine is called biological naturalism," and I invented it on the spot and now I'm stuck with it. That was on the earlier slide. But the idea is that the right level for explaining consciousness is biological. But it's just part of nature, we're not postulating anything supernatural or anything that stands outside of nature.

Okay, so that I think is our first mistake that we need to overcome, and that's the biggest one. If we could get out of our tradition that says that these are mutually exclusive, then a lot of the intellectual obstacles to getting a naturalistic account of consciousness would be removed. Notice that I'm not now defending materialism, which says really consciousness as such doesn't exist, you've got to reduce it to something else or show that it was just an illusion. And I'm certainly not defending dualism that says no, no, they're really two different realms that we live in, the mental and the physical.

Now a second mistake we've got connects with a notion of causation. And we're inclined to think that whenever A causes B, they have to be completely different phenomena, different events, that causation is always a relation between discrete events ordered in time. I want to say there are other kinds of causation than that. If you ask yourself what's the causal explanation of the fact that this rostrum exerts pressure on the floor the answer is given by gravity, there's a constant gravitational attraction. But of course gravity is not the name of an event, it's the name of a permanent fixture of the universe. And if you ask what's the causal explanation of the liquid behavior of this water I have here, I could tell you a story about how the molecules are rolling around on each other in a more or less random fashion, and that the behavior at the molecular level causally explains the behavior at the system level. Similarly the behavior at the molecular level of the molecules moving in vibratory movements in lattice structures explains the solidity of the rostrum or the solidity of the floor. I want to call your attention to that, because that is bottom-
up explanation, where you explain a surface feature not by citing a preexisting event, but rather by citing the behavior of the microelements of which the system that has the surface feature is composed. And I think that is the right model for seeing the relation of neuronal processes to consciousness. Consciousness is caused by the neuronal processes, but the form of causation is bottom-up causation, so you can have the higher-level system feature causally explained by the lower-level behavior of the elements, even though the system is made up entirely of those elements.

I noticed some of the speakers were reluctant to say cause. One of the ways we have a fudgingness is we say, "Well the brain gives rise to consciousness." That's a hedge. Let's come right out and say it: the brain does it, the brain causes it. It isn't something that sort of just squirts out in a sort of vague way. No, it's really going on as a real feature causally explained by the brain.

Okay, reduction. Well, how much time have I got? I mean I'm just . . . have I got another ten minutes? Who's the boss? Okay, all right, all right, here we go. Reduction. The problem with reduction is that it's almost a religious term. A lot of neurobiologists tell me science is reductionist. They say that with tremendous sense of achievement. I learned it in graduate school is the idea. And a lot of people object to me by saying, "Too reductionist, too reductionist, this whole account."

And so I went and looked at the literature, and the truth is nobody knows what they mean by reduction, or rather there are half a dozen different things that they mean. There's logical reductions and causal reductions and ontological reductions and eliminative reductions, and I'm just going to tell you a couple of distinctions you need to keep in mind. First of all, what we mostly want in the sciences is a causal explanation. Now typically when we get a causal explanation, we will make an ontological reduction on the basis of the causal reduction. Now what does that mean? Well if you know that the behavior, the liquid behavior, is causally explained by the behavior of the molecules, then you have causally reduced liquidity to molecular behavior. Analogously, if you know that the solidity is causally explained by the behavior of the molecules, then you have a causal reduction. A causal reduction tells you that the reduced entity can be causally explained by the behavior of some other entity, some lower-level entity. Now typically when we make a causal reduction like that, we make an ontological reduction. We then say, "Well, liquidity just is the behavior of the molecules, or solidity just is the vibratory movement of the molecules in lattice structures, or the color red just is a certain type of wavelength." The causal reduction typically leads to an ontological reduction. But we're reluctant to do that with consciousness. Why? I mean suppose we had a complete causal account of consciousness, we could say exactly what was causing consciousness in the brain. We would still be reluctant to make an ontological reduction, to say consciousness in nothing but . . . and then follows your favorite theory, it's neuron firings at the rate of 40 Hertz synchronized between layers 4 and 6 of the cortex and the thalamus, let's say. I mean that was
one theory. It didn't work, but anyway that's the kind of theory, something like that has to work. Now we would be reluctant to make an ontological reduction, to say, "Well that's all there is to consciousness," in a way that we're not reluctant to make an ontological reduction with solidity and liquidity. What's the difference? Well I think it's not that there's some deep metaphysical difference. I mean solidity continues to feel a certain way and liquidity continues to behave in a certain way. The difference is that in the case of consciousness, the whole point of having the concept of consciousness is to name this qualitative, subjective sequence of experiences. So even if we got a complete causal account, we would still be reluctant to go the next step and say, "Well that's all it is, it's really nothing but that," because of course it's our life, we actually live this sequence of qualitative experiences. So I think we will get a causal reduction of consciousness if we have an ideal neuroscience. That's the aim is to get a complete causal account. But the causal reduction of consciousness will not lead to an ontological reduction in the way that causal reductions typically do lead to ontological reductions because we'd lose the point of having the concept if we made the ontological reduction. We still need a vocabulary to describe these qualitative subjective features of our sentient and aware life.

Now there's another distinction you need to make in reduction, and that is between those reductions that get rid of the reduced phenomenon where you show it was an illusion, there wasn't anything there, and those that don't get rid of it but just explain what it's made of, what its basis is. So I guess we can do an eliminative reduction on sunsets. I mean sunsets are just an illusion. The Sun does not really set over Mount Tamalpais, it appears to from my house, but it doesn't really, that's just an illusion. But we don't get rid of objects by saying, "Well really they can be reduced to molecules." That's a non-eliminative reduction. Now the question that a lot of people ask is, "Well if we can do an eliminative reduction of sunsets, or rainbows, let's say, why can't we do an eliminative reduction of consciousness? Why can't we show the whole thing was just an illusion?" And the answer is this: the elimination . . . eliminative reductions rest on the distinction between reality and illusion. So you can do an eliminative reduction of the sunset because it's an illusion, an eliminative reduction of the rainbow because it's an illusion. But the interesting thing about consciousness is [this]: where consciousness is concerned, the illusion is the reality. That is, if I now have the conscious illusion that I'm conscious, then I am conscious. The way that our traditional eliminative reductions work, by showing a distinction between appearance and reality, won't work for consciousness because the appearance is the reality. That is, if it consciously seems to you that you're conscious, you don't have to worry, you are conscious. If some guy comes to you and said, "Look, we've done a study and we've shown that people who meet your profile are in fact zombies, you have no consciousness whatever," you don't have to worry, you don't feel, "Oh, you know, maybe those guys are right." Don't be intimidated.

All right, so I resist the temptation to talk some more about reduction. I've talked longer than I meant to, so I'm going to skip identity. Okay, identity is another
philosopher's favorite, and I'm going to go [to] my last slide, as lecturers always announce and everybody feels relieved. Unfortunately it's got a helluva lot of stuff on it, and it was really because of that that I asked for this paper to be handed out. And I didn't even ask for the slide but now we got it. Okay.

**Traditional Distinctions Between Mental and Physical**

Now I just decided for fun, let's make a list of our tradition. What is it that makes the mental so different from the physical? Why can't we just face the obvious fact that the mind is an ordinary part of the biological world like digestion or photosynthesis or the secretion of bile, it's just a normal part of our life? And the answer is we've got this tradition that says they're radically different. Mental is subjective, physical is objective. The mental is qualitative, the physical is quantitative. The mental has a first-person mode of existence. That means it only exists insofar as it's experienced by some eye, some self, that has it; whereas the physical has this third person or objective mode of existence. It exists apart from anybody's thoughts and feelings. Furthermore, the mental has intentionality. That's a fancy word that means aboutness or the directedness. My mental states can be about other things, so I can now be thinking about George W. Bush even though he's in Washington and I'm in New York, and physical things don't have that feature. I mean words can do that, but that's only because we use them, we employ intentionality on them. So the physical is non-intentional, and then we have this tradition that says that mental states—we get this from Descartes—are not spatially located, and they're not extended in space, whereas everything physical is spatially located and in general spatially extended. We think that the mental is not explainable by physical processes, where we think the physical world has to be causally explainable by microphysics. And then of course, we think of the mental world, at least some of us do, as somehow incapable of acting causally on the physical because they're in these two different realms, whereas the physical world we think is a causally-closed system. If the mental really existed, it couldn't affect the physical world because the physical universe is causally closed and nothing nonphysical can ever come into it.

Now I want to say that chart embodies some of the most absolutely fundamental mistakes of our civilization, and I want to chop it right in half right now. And it's this: I want to say start with the mental and grant all those stuff, that it really is subjective, I really do have these qualitative, subjective feelings. They have a first-person mode of existence and they have intentionality. But now scrap the bottom, and move the mental . . . I've even got one of these weapons that they all use, here we go, watch this. Well, I want you to notice . . . there we go, all right. Scrap it right here, move all of this over to here and just say the physical world happens to have features that are both ontologically objective and subjective. Some of them have a qualitative feel to it and a first-person mode of existence and intentionality. They're products of certain neurobiological processes. But notice, as we've been seeing for the past two days, they're spatially located and spatially extended. You don't believe it, we'll turn on our fMRI and I'll show you exactly where they are, and
we're going to find out more, we've even got single-cell recordings. Furthermore, we're going to causally explain them, as we causally explain everything else, by showing that they're the result of bottom-up forms of causal explanation, and we're going to show how they act causally. And I want to say a little bit about that.

I said, with perhaps too much self-confidence, that my decision to raise my arm causes my arm to go up, but of course there is a story to be told about the activation of the motor cortex. We know the neurotransmitter, it's acetylcholine. We know how it goes to the axon end plates of the motor neurons, and there's a whole long story to be told about the cytoplasm of the muscle fiber and the actin filaments and the myosin filaments. Now why isn't that story enough? That is, you know a microstory, why doesn't that tell a story and the conscious decision to raise your arm just goes along for a kind of free ride, it doesn't really do any work. There's a name of that, it's called epiphenomenalism. That says the mind is there but it doesn't do anything, it's like a froth on a wave, it doesn't perform any work. I want to say the way we should think of it is like the car engine. It's true that there is a story to be told about the oxidization of the hydrocarbon molecules and the impact of them on the metal-alloy molecules of the piston, but when you go to your car mechanic, you don't talk about that, you say, "The damn thing won't start." What you don't say is "Look the passage of electrons between the electrodes is insufficient to oxidize the hydrocarbon molecules to the extent that the oxidization becomes self-sufficient." You just say, "The damn thing won't start, I think it's in the plugs."

Now the point is it's not that there are two different domains being described; it's the same domain from beginning to end, just different levels of description. In Berkeley you've got a lot of unemployed physicists working as car mechanics, and you might tell them that story about the oxidization of the hydrocarbons, but for normal human mechanics, you don't have to go to the micro level. And it isn't that there's some metaphysical problem about how can the spark plug ever work when really the only thing doing the job is the passage of the electrons between the electrodes; it's the same system being described at two different levels. And I want to say that's how it is when you raise your arm, it's the same system being described at different levels.

All right. Well let me just summarize, and I've only really begun this talk, but I think I got across the main message I want to get across, and that is if we can overcome certain traditional errors, if we can overcome the mistake of supposing that we live in two different realms, then we can accept consciousness on its own terms, and begin to investigate, and now the investigation is well underway, how exactly it works in the brain. There are two different research proposals. I sort of hope the other guys win because it's an easier research project, that is, what I call the building-block approach. I think the unified-field approach is pretty tough—you've got to figure out how massive rates of neuron firings in big chunks of the brain, presumably the thalamocortical system, how they cause the system to be consciousness, and that seems to me a much harder research project.
But I'm delighted with the way this conference has gone, and as several speakers have remarked, it would have been unthinkable 25 years ago. I remember when I first got interested in this, I thought, "Well, you know, I'll go talk to these medical scientists in San Francisco," and there are a lot of people in medical school interested in the brain, and I went over and I talked to a famous neuroscientist, "Well why don't you guys get busy and solve the problem of consciousness?" And his answer after much discussion was this, he said, "Look, it's okay in my discipline to be interested in consciousness, but get tenure first, get tenure first." And I think if I had to describe the intellectual revolution that has gone on, and this is the kind of thing that drives intellectual revolutions, you can now do serious work on consciousness without tenure.

Thank you very much.

**Question and Answer**

**David Cohen:** Just one question, because we're running very late.

**Man:** Thanks. You have to get up fast if you want to ask questions here. You've given a very excellent but also very loose definition of consciousness, and throughout it I was struck by some of the apparent similarities between the way you were describing consciousness and things you might describe in machines and computers. For instance, you referred to the guy who wakes up in the dark room doesn't have any awareness, but he's nonetheless active. I turn my laptop on, and although there are no programs running, the sucker's on. Given that and certain other comments, I was wondering if you could comment first on the possibilities of true consciousness arising in machines. And then second—and this is perhaps more in relation to some of the researchers who've been up here—what might we eventually be able to learn about our own consciousness by sort of watching this accelerated evolution of intelligence in machines?

**John Searle:** Okay, could everybody hear the question? The question was, What about the prospect of machine consciousness, and how much can we learn from the progress of machine intelligence?

Now along with the dualism of the mind and the body, it seems to me another colossal mistake we make, and again it goes back to the seventeenth century, is this opposition between humans and nature and between humans and machines. If by *machine* you mean a physical system capable of performing certain functions, then we are machines. It seems to me there's no question we are machines, and that we've got these submachines, like our heart and our liver and so on. So there isn't a question about machine consciousness, I'm it and so are you.

And now the question is, Well how about an artificial machine, couldn't we build a conscious machine? Now I think we ought to hear that question like the question
can you build an artificial heart that does what the heart does? And we know the answer to that. The way we got the answer to it was by figuring out how the heart works, and then building an artifact that would do the same thing. We don't know how the brain works, we don't know how the brain produces consciousness, so until that, we don't know how to build an artifact, an artificial conscious machine. Now the problem with the notion of machine intelligence is that the machines we're talking about aren't actually intelligent—that's just a metaphor, that's an observer-relative ascription that we make to them. You see, when I do addition, I'm not very good at it but I can do three plus five equals eight. Now when my pocket calculator, I punch in three plus five and it doesn't think, "Ah, that's a three and that's a five, I've got to print out an eight," it doesn't think that at all because it doesn't think anything. It's just a hunk of junk. It's just an electronic circuit that we've programmed. Any intelligence in the computer or in the pocket calculator is entirely in the eye of the beholder, it is entirely observer-relative.

When Deep Blue beat Gary Kasparov, I was besieged by reporters. Fortunately I was in Europe and they weren't willing to spend much money on long distance, but they wanted to know isn't this a blow to human dignity, doesn't this show that the machines are really taking over? You know, you can imagine the questions that one would be asked. But the answer is, of course, it's no more a blow to human dignity than the fact that any pocket calculator can outperform any mathematician in the world. We've designed this hunk of junk to do this kind of stuff and it's terrific, but nobody should think it's of any psychological relevance. Deep Blue didn't beat Gary Kasparov in any ordinary sense because it didn't play chess, it didn't know that it was winning or losing, it didn't know that this was a pawn or this was a knight. I actually did some research on this and found out how it worked, and it is a terrific technological achievement. And as you all know, the problem in chess is the exponential problem, where you just get too many things. Well, Deep Blue could calculate 200 million moves in a fraction of a second. And that's of no psychological relevance. But Deep Blue doesn't know that it's playing chess, it doesn't know that this is a chess game, it doesn't even know that these are numbers. It's just a fancy electronic circuit without any consciousness or mental life at all. And that's a general model for so-called machine intelligence.

Now what's wrong with computers? Well the answer is—and this is what a lot of people don't get—it isn't that the computer is too much of a machine to be conscious, it's not enough of a machine. Because you see, our brain really is a machine, its operations are defined in terms of energy transfer. Computation doesn't name a machine process, it names an abstract, formal, model-theoretic, algorithmic process that we've found ways to implement in machines. But computation is purely abstract, and we put this by saying it's all syntactical, and the syntax by itself, the zeroes and ones by themselves, don't carry any causal powers. We implement that in machines, so the thing you buy in the store is a machine, but the actual computational processes are not machine processes. What's going on in here, however, are actual machine processes.
Brain and Mind
May 14, 2004

Nancy Wexler, PhD
Concluding Remarks

Introduction by David Cohen

David Cohen: I think we're going to need to move on, we're really quite far behind, and I want to express my real gratitude to the speakers in this morning's session. It really has been a remarkable session.

And I am also going to forgo my brief closing remarks because I think the spirit of them was so entirely captured by John Searle just a few moments ago, and I think it's well summarized that one can study consciousness before tenure. It's now my pleasure to introduce Dr. Nancy Wexler, who will offer closing remarks on the overall symposium, and I won't embarrass Nancy by telling you how long I've known her. Nancy received her AB from Radcliffe and her PhD in psychology from the University of Michigan, and she's currently Higgins Professor of Neuropsychology in the departments of neurology and psychiatry at P&S, here at Columbia. She also serves as president of the Hereditary Disease Foundation.

Nancy has had a rather remarkable and uncommon career. Beyond her widely known and broadly ranging research on Huntington's disease, she has contributed in substantial ways to public health policy, genetic counseling, federal health administration, and so on. She currently holds or has held numerous public policy positions, including chair of the joint NIH-DOE ethical, legal, and social-issues working group of the National Center for the Human Genome Research, and serves on the board of directors of the American Association for the Advancement of Science, and so on. Activities such as these earned her the prestigious Lasker Public Service award in 1993, and indeed her awards are numerous including several honorary degrees, election to the Institute of Medicine of the National Academy, election as a fellow of the Royal College of Physicians, member of the European Academy of Arts and Sciences, and so on.

Again, I'm most pleased to present to you Dr. Nancy Wexler, a long-standing friend and colleague.
The Dialectic Between Basic and Clinical Science

Nancy Wexler: You're amazing to still be here and if you start leaving I won't take it personally, I know we're way over time. And yesterday I said to Gerry and Tom Jessell, "I can't possibly summarize what's going on, this is like capturing an aurora borealis, so is it okay if I just give some themes that I've heard?" And they said, Fine, and you can even talk about what you want, which is a dangerous permission.

Gerry opened the day yesterday morning with an incredible [inaudible] to the changes that have happened in the last 250 years, but also with this very intimate relationship between the critical nature of basic science as its understanding for human diseases. And that he listened to every single one of the speakers, every person talked about how basic science and disease coexist. The basic scientists give the people researching disease avenues to look for, and even this morning people were talking about how you understand consciousness. John Searle was talking about split brain, Christof was talking about using electrodes during epilepsy surgery to understand the nature of consciousness, Nancy was talking about how a stroke affects where you actually see pictures and understand faces, so there's this unbelievably kind of intimate conversation, and I think it's so critical because we have at Columbia all of this kind of interdisciplinary conversation going on. We have more patients just across the street and up the way, we have more scientists, and have more interaction, and we have spectacular meetings like this where everybody can hang out and talk to each other, and I think that's really how we're going to advance the next 250 years.

So let me say a few other things: I'm actually the only person here who's actually using a slide. And again I want to go back to something that Gerry Fischbach was talking about, this intimate dialectic between basic and clinical science. If you can remember, yesterday morning Gerry was talking about snakes. Yesterday morning was the morning of the snakes. And there's a particular snake in Indonesia that was making a toxin that was causing a major public health problem. So when that *Bungaris multicinctus* toxin was injected into rabbits, the rabbits ended up with huge flapping ears. Now the danger in a situation like that is that students working in the lab who are basic scientists are going to walk in, see those rabbits, and throw them out, because they're going to think this is a failure of the experimental preparation. You know, let's do it on iguanas, or let's do it on something which doesn't have ears so that, you know, we don't have to have this problem of having this thing flopped in the bottom of the cage. The fact that a person looking at those floppy ears could say, "Gosh, this looks likes myasthenia gravis" was amazing, and, in fact, I think it was a graduate student or a technician. But you have to be able to make the connection to the human disease to know what the basic science is related to, so you really have to have that intimate knowledge at all times.

And this is a family that we've been working with in Venezuela who has Huntington's disease, and these types of diseases again are always . . . have this
kind of dialectic of the basic and clinical understanding. Now part of the afternoon
and the morning of the snake yesterday—actually, you can't really see snakes, but
there's a kind of jungle back there. And there are an awful lot of snakes in the
jungle, and this is why families have decided to move onto stilt villages, and it
really is because there's too many snakes back there for comfort. We never go
back there. Sometimes you can see them swimming around in the water. But one
of the things that happens is that as a disease develops, people fall into the water
and drown. Now that's an environmental modifier of a gene, after a fashion.

Symposium Summary

And another theme of these talks is plasticity, the effects of genes, and the effects
of modifiers that effect of environment, and how do you find all of these and how
don't you put them together in a way that makes sense? Now after Gerry's
presentation, Rod MacKinnon gave us an elegant portrayal of potassium channel
in action, which he has crystallized. And it was quite extraordinary because you
really could see how the channel opened and closed and how the molecules
danced together. He also just completely, not talking to anybody else, had a snake
up there, because he was talking about the effects of toxins and how these toxins
from the green mamba, from the cobra, from the tarantula, from the sea anemone,
scorpion, and wasp can enter the calcium channel and essentially kill the creature
or paralyze the creature. So it's this kind of—again another theme running through
these talks has been the conservation of species and the conservation of
molecular material. So that all the way from a scorpion in Chile, as he showed us,
to the most primitive creatures living in thermal vents in Japan, they have the same
DNA, they have the same structures, and they have the same kinds of . . . one had
a toxin that paralyzed, but it was very similar to the other structure. So it's
extraordinary how these kinds of use of the past and this sort of looking back and
forth, and back and forth across different animals and different people and different
diseases leads you to some common phenomena.

Richard Axel had a snake as an example of a faculty actually, which humans don't
have, along with bats the ability to do sonography or hear, you know, sonar. And
Richard also talked about the conservation of species, but in the smell system that
he was talking about he actually came to a rather different conclusion than Rod
MacKinnon. I guess I should stand over here, you're getting two of me, right? So
Richard talked about, Richard had a new version of sense and sensibility, which
was quite fascinating. And he talked about the fact that there are 32 genes, as he
called promiscuous genes, for the senses of taste and vision, but there are nine
hundred genes for olfaction, and these are chaste genes, according to Richard,
because you need each one of them and they serve a different function. He also
pointed out that the mammalian brain and the Drosophila brain have the same kind
of arrangements in dealing with olfaction, but that they arrived at it independently.
Which is quite different than Rod's notion of having conservation of the same, if
nature got it once billions of years ago why don't we just keep it and drag it along
and tinker around the edges? What Richard was talking about was really
something completely fundamentally different; which is, nature has to do it twice independently but she ends up doing the same thing because there's probably not that many different ways of getting it right, and why not? So you end up at the same place and the same kind of structure in the fly, and in Gwynth Paltrow, and in Richard Axel, and all of us here, but we got there independently. So you need even plasticity in thinking about these kinds of systems to understand them.

Tom Jessell talked about how the whole nervous system evolved and had kind of a vocabulary of words that spanned *Drosophila* language, sonic hedgehog, Hox genes, ETS, cadherins, and the whole system again elegantly described to as, How sensory neurons are innervating a muscle, how they allow—we heard one explanation of how your arm went up and down, but Tom Jessell could've explained it in quite a different way in terms of the genetic biology of it. And again we know that devastating failures of those systems can lead to diseases like ALS and other ones, Lou Gehrig's disease. And Lou Gehrig was actually a student here.

We heard from Huda Zoghbi, who really I think epitomizes the kind of modern molecular crusader, because many of the talks talked about plasticity of the nervous system. But the speakers themselves, the investigators, are very plastic in what tools they used and how they wanted to go about learning how to understand their diseases, and Huda was one of the first discoverers of spinal cerebellar ataxias and causes, increased repeats like Huntington's and some of the other ones. [She] made a mouse with that gene in it, and is working on therapies for those diseases, but at the same time, an extremely different disease. She was looking at these little girls with Rett syndrome who came in to see her, and it was just devastating because these were healthy, beautiful, gorgeous little girls, and then after the age of 3 or 4 they would suddenly reverse development and begin to die before her eyes. And Huda felt very helpless to stop this, so she thought, let's try to find the gene. Now she used four families, which were practically nothing in terms of big families, and she also noticed that these are just little girls getting sick, and not usually boys. So as a geneticist and a clinician she says this has to be on the X chromosome, she started looking on the X chromosome, and bingo, after many, many years, they found a gene on the X chromosome called MeCP2, which controls how a gene is transcribed, whether it's expressed or not expressed. And then immediately Huda took that gene, made a mouse model of it so she could try to do treatment—she had to have a mammalian system but mice are too slow, so she put it in a fly and she actually got the Rett syndrome into a fly eye, so you could see the eyes dying in front of you, you could treat it right in the eye. She also put it in the whole nervous system and you could see the wings just shriveling up and dying. So immediately, because the *Drosophila* genetics are so powerful, Huda is trying a lot of different genes to see what makes the disease worse and what makes it better. Whatever she tries in the fly she tries immediately in the mice, and whatever works there the hope is to go into humans. Now we're all very grateful to Thomas Hunt Morgan who was right here actually working out the fly genetics, so we have it ahead of time.
The other speakers yesterday afternoon were struggling because they didn't have genes to work on, and a lot of their talks were about the search for these genes and how frustrating in a lot of these kind of complicated multifactorial psychiatric diseases to actually find genes. So Judy Rapoport, Sir Michael Rutter, Nora Volkow yesterday, all of whom used, again, extremely novel technologies, were somewhat frustrated by what's happening with the gene searches. And I'm sure that those are just temporary pauses, but it does show you the difference between being able to make a model system if you have a gene and sort of being stuck with looking for a gene. Judy had unbelievable pictures of brains dying, the cells dying in living purple color, which is frightening, and she also showed how these same kinds of patterns of death can be true in normal systems. And that was kind of alarming I think to all of us, because you say wait a minute, but it was a very good point because there's kind of this slow slide between what is normal and what is abnormal. And this point was really driven home even more by Mike Rutter, who showed how just problems with reading could actually lead to all these other concatenations of difficulties. And he also talked about the genes for autism, for attention deficit disorder, for dyslexia, and how many of these disorders disproportionately affect males. So that's a clue, and hopefully like the Rett syndrome that might lead to us to some genes.

So these diseases really affect you dramatically, and they affect the family and they affect the next generations looking at these. Many people talked about using twins and compared monozygotic and dizygotic to get a handle on what's genetic. Nora Volkow, in talking about addiction, again looked at dopamine receptors to see whether or not having more receptors is protective or makes you more vulnerable. Almost everybody has used imaging on both days to look inside the brain.

**Applying Knowledge to Drug Development**

Now I just want to super-briefly talk about what's going to happen when we have all of this knowledge, because there's almost no new drugs being made, there's just a paucity of development of new drugs. And the most critical thing about this chart, I don't know if you can see it very well, but the top lines are the amount of money that's being spent worldwide in research and development in sales for all new drugs worldwide, and then the bottom line going down is new drug launches. And even if you can't see the absolute numbers, the number is just declining dramatically. Dramatically. So even though there's more money being spent in research and development and there's more sales so there's money coming in, it's not being translated and having any of what we're talking about and what we're doing translated into new therapies. And even more frightening, the drugs for the nervous system as a total proportion of the pie is decreasing. So the biggest one in there is cancer drugs, and the cancer drugs are assuming a larger proportion of the pie. But infectious diseases, which is AIDS and [inaudible], is really being squeezed.
Now I think all of us have heard and talked about the problem of just the cost of new drug development. And again these are statistics that cover worldwide markets. They are about 880 million dollars or so for just developing a new drug. And we have got to be able to do something about these costs, particularly with all of our DNA on the Internet, every single one of us has a genetic disease and an orphan disease, because each one of us with a greater specificity of pharmacogenomics, we're all going to be having an orphan drug and uninsurable.

And the last figure is the number of new drug launches. In 2002, there were thirty new drugs; that’s it, worldwide that were launched, that were out there. I mean just listen to the number of diseases we talked about today, thirty. And the numbers are getting smaller.

So where are we? What do we do to try to reverse this trend? Because it would be horrendous if all of the kind of phenomenal wealth we heard about the last two days cannot be translated for the benefit of people who are suffering. And we need economics, we need persuasion, we need politics, and we need everybody here today to try to reverse this trend and make these drugs more accessible.

Thank you very much, and enjoy the rest of the day.