Christof Koch:

Thank you very much Martha for that charming introduction. Thank you John and thank you Sheila for coming here, I guess we’re neighbors, she’s a professor at Pomona, and I am just twenty miles down the road from there at Caltech. So, let me without further ado talk about my life; so, I’ll talk about the biology of consciousness. If you want more information you can just go to my website or you can google the research I’m going to talk about. I read in preparation for this lecture, I read the book that Benjamin Pinkel wrote on this topic, and, essentially, he is an engineer and a physicist, and I am also by training, as you heard, a physicist. I’m, in fact, both a professor in both engineering and in biology at Caltech. So, I think about consciousness and the mind-body problem both from a physical point of view as well as from a biological point of view, as well as from a point of view that we all have, which is really where consciousness meets the road, where the rubber meets the road, is, all of us are conscious, so, with René Descartes, there is one thing that we can conclude, we can be extremely skeptic, as René Descartes was back in the early seventeenth century, we can doubt everything, we can doubt that I have a body, I can doubt that the world actually appears the way it appears to me, but what I cannot doubt, what is not subject to any further doubt, is the fact that I am, that I can feel things, I can feel pain and pleasure, and I can feel happiness and I can feel joy, and I can be depressed, and I can see things, and I can hear things, and I can smell things.

It turns out, as a hundred years of psychology and now neuroscience has shown, there are many, many things in your brain, there are many things in your body, in fact, most of the things in your body and brain that go on without you feeling anything. And so, the wonder of it all is how is it that we have this thing called consciousness? It would be totally bizarre if you described it to somebody unless you had it yourself. That is one
of the problems with consciousness; you can’t even define it without if you go, for example, to the Oxford Standard Dictionary, you see consciousness is defined usually in self-referential terms. It refers to itself, and people spend PhD theses trying to define consciousness very rigorously and it can’t be done, certainly not right now. So, like any good scientist, I don’t worry too much about rigorous definition; it almost never happens in science. Definitions are something later on for the textbooks once a science has, sort of, matured and you really understand it. I’m going to talk about the state that you find yourself in right now or that you might find yourself when you are driving, or watching a movie, or climbing, or running, or doing all the other things we do, or that an animal finds itself in, and what is the relationship between the body, between the physics of it, and the subjects of states? And how can we manipulate, and what can modern science do about it? This is really the topic of Benjamin’s book.

So, I should say that all the conceptual work of this I’ve done over the past two decades, really, with Francis Crick, who, when he moved from Cambridge, England, he stopped doing work in molecular biology and, rather than settling down to a well-deserved retirement, he came to Southern California and started to think about consciousness. We’ve got many, many topics, we’ve got many papers on this topic, so the conceptual work is really joint work with Francis. (Can we have the lights down? Is it possible to dim the lights? I can do that? I can?) Alright, while we do that...So, you should all be having a particular conscious state now—if you attend to me and not are thinking of lunch or something else—you should all be (thank you John), you should all be seeing red. And, there is a particular way it feels red. Philosophers like to talk either about the consciousness of color or consciousness about pain - probably tells us something about philosophers that those are the two favorite things they choose. So, here we all see red. Now, I am not saying that we all see the red exactly the same, we know we don’t, for example, many of you women will have two alleles, will have both of the alleles for the long wavelengths for the receptors, so you will see, you will be able to see shades of red that guys like me who only are trichromat, at best, and some of the
men are, of course, are only dichromat, can never see. But, we will roughly see all the same.

The fact is that we see this, and it reminds you of other red, and it reminds you of sunsets or blood or things like that. Philosophers talk, here refer to, this ineffable aspect, this “feeling” part of red as qualia. The consciousness is made up of different, the elements of consciousness are qualia, and you can have a...(the singular of qualia is quale), you have one quale for color, you have a quale for being angry, you have a quale for having a particular type of toothache, those are all different aspects of consciousness. What is common to them all, so, for example, here they are...what is common to all these five different percepts, is that they all involve the quale of red. And the question is how can a brain generate a quale, how can the brain generate any quale?

I first thought about it very vividly many years ago when I was teaching a course in Woods Hole in the Maths and Computation in Science course. I had a bad toothache and I was lying in bed and I was taking aspirin, and it still hurt. And then trying to distract yourself, you think why does it actually hurt? Which seems like a trivial question, but it is actually one of the deepest questions you can ask. The simple answer is, well, it hurts because your tooth is inflamed, and then, so, you have neural activity, you have electric action potentials that go along here to the [caminus] nerve, go inside here, go back inside the spinal cord, a switch over at the thalamus and end somewhere in cortex proper, and they give rise to electrical activity, and that is why it hurts. That’s what the doctor says or that’s what the neuroscientist says, but that’s really no really explanation at all, because that doesn’t tell me why it hurts. All I can see is what happens, you have a tooth inflamed here and then some neural activity over here. The neural activity amounts to iron sloshing around, there is sodium, potassium, calcium, other ions that slosh around in the brain, but why should they hurt? I mean, it’s just a movement of ions. Somewhere over here, if those ions move around, I think about a
flower, or, somewhere else, I think about pain. But, in each case, there is nothing different if you look at the ions themselves, the ions here and here, I can assure you, look exactly the same, and the electrical activity over here and here, they’re action potential, they are little pulses of electrical electricity, they don’t look different. Yet, over here, it’s pleasure, and over here, it’s pain, and over here, it’s seeing red. Now if I take this gorgeous computer, it’s OK, it’s a Macintosh Air, the best computer there is, you take a thermometer, you couple it into your Macintosh, and let’s say you program it so that if the temperature exceeds, let’s say, hundred twenty degrees, then it prints out in letters “pain.” So, you go into this room, the room heats up, and pain comes out. But, then, nobody would think that the computer actually is in a state of pain, you would say, “Well, that’s just electrons switching, ultimately, there is a bunch of charge of electrons flowing onto a transistor gate, but that’s not pain, that’s just physical stuff.” Same thing if I look at your brain when you have pain, I can say, “Well, I don’t know about pain, all I can tell is that there is electrical activity in your brain, but I don’t see where the pain comes in.” But the fact of the matter is we have these conscious states, and they are very mysterious, and we don’t understand how they arise. So, this is the heart of the mind-body problem as first conceived by Aristotle two thousand three hundred years ago.

Now, today, what we can do, what Aristotle couldn’t do, we can look at inside the brain while you’re actually having a state of consciousness. So, folks, here it is fictitious, it is my brain, it’s a big brain, it’s an egghead brain to be precise, and, here, I’ve painted on artificially the neural activities supposedly if I am looking at red. Now, this is called a third-person account of consciousness. I can look at this brain from the outside—you can do it, I can do it myself—I can look at my own brain and I can observe, “hmm,” whenever this brain sees red, then, for example, your activity in this part of the brain, but, of course, what you can never get is a first-person account, only you, only the conscious entity can have a first-person account. What makes consciousness so intractable and so much more difficult, than, for example, trying to synthesize the noble
biological organisms, or trying to understand superstring theory, is that both of those are difficult and hard problems, but they don’t have a first-person perspective to them.

Superstring, you don’t ask, “What does it feel like to be a Superstring?” You just like to have a satisfactory physical description of Superstring. You are trying to test the prediction of the theory. But here, what a theory of consciousness, in the fullness of time has to explain, how is it that certain systems - my brain, for example, or your brain - how is it that these certain systems can have subjective states associated with very particular types of activity in particular parts of the body? So what you should do here is fixate, look at, for example, the “B.” Look at the “B” and then tell me what you see. You should all see a blue sphere, right? Yes? No? Hello? Alright, good. And, you should also...keep on fixing down here, you should also see yellow squares, right? Otherwise, just come see me afterwards. What you might notice if you fixate here, do you notice anything funny? Yah, exactly, thank you. So, did you all see that? Sometimes one, sometimes both, squares disappear. It’s a nice illusion invented by these people, and the point that I am showing it is that, to come back to the definition, because many people will say, “Wait a minute, Professor Koch, before we can study consciousness we have to define it.” And, I already made the point that science usually in practice never proceeds that way. So, but here, you have, sometimes, you have consciousness for yellow and sometimes you have no consciousness for yellow. When you don’t see it, it is just not there. When you see it, you have a vivid sense of yellow; it reminds you of, I don’t know, of the sunset, of the sun, or maybe of Van Gogh’s “Sunflowers,” whatever else.

So, where is the difference in your brain between the state when you are conscious of yellow and when you don’t? That’s the question we are now trying to ask, and scientists are trying to ask, the difference between the states of your brain when you see it and when you don’t see it. Now, before we do that, we have to make an important conceptual distinction, so, there are two very different ways of talking about
consciousness. There are states of consciousness and there are conscious states. So, states of consciousness is traditionally referred to in a clinical context, for example, in a bad car accident, and you may be in a coma or you may be, for example, if you remember, Terri Schiavo, remember this unfortunate lady in Florida who died after being fifteen years in a state that as far as we can tell, as far as science could tell, she had no sensation whatsoever? There was no consciousness; there was no Terri Schiavo in there. She was, from a practical point of view, it was flat; it was like she didn’t have any consciousness as far as we could tell, yet, she was, of course, alive. And, so, in her case, the basic enabling conditions weren’t met because of the anoxia that she suffered fifteen years earlier. Her heart stopped briefly. Parts of the brain were destroyed, particularly those parts in the brain that are necessary for you to have consciousness at all.

They’re parts of the brain called the mid-brain, there’s a collection of nuclei known as the mid-brain reticular formation. It’s a very complicated neurochemical set of different nuclei, and there is also part of the thalamus, in the middle of the brain, called the mid-line part of the thalamus, and they are really essential for you to have any consciousness whatsoever. You can have big strokes and big lesions in your cortex and you will have very specific deficits, but if you have tiny lesions in these parts of the thalamus of the brain stem, you may not be conscious at all. You may be like Terri Schiavo or like other cases—many of them. In the US, there is estimated to be fifty thousand people to be in a state similar to her called PVS, Persistent Vegetative Syndrome. And, so, we know quite well the conditions that need to be present for you to be conscious at all. What philosophers and what, I’m personally, most scientists are interested in is not these general enabling states, although they are, of course, critical from a clinical point of view. We’re interested, given that you are like me and like all of you right now, where does the content of the specific conscious sensations, how does that arise? So, the fact that you see yellow or you didn’t see yellow, in both cases, you were conscious, in general, but one point, you had this content yellow and, in the next
second, you saw something else. Or, suddenly, if there is a lot of noise, the content of your consciousness will shift and attend towards the noise.

So, the mystery is, how does this specific content arise inside your brain? Alright, as scientists we look at the continuity of our species, because we are just one animal among many other animals; we are the dominant species right now, but that may change. So, these are dogs, these are my dogs, I love dogs, they’re gorgeous, and, for those of you who have dogs and cats and other animals, there is probably little question that these guys are conscious. Now, I’m not saying that the consciousness of animals is the same as our consciousness, particularly what these guys, dogs among others, but most animals lack, they don’t have inside; they don’t have self-reflection. There is some evidence in some of the great apes for some minimal amount of self-consciousness, but, in general, leaving those aside, a dog doesn’t think and reflect upon himself. A dog has pain, a dog has pleasure, a dog can be very excited, a dog can be depressed, it can be in pain, all of those things. The belief is, certainly among biologists, is that we share consciousness, certainly with most mammals, well, we don’t actually know how far we share consciousness outside of the mammalian kingdom. What we can say, why do we think that animals are conscious?

Well, for one, the behavior is very similar; so, if your dog is in pain, you know it, even though the dog can’t talk [to] you. It’s whining, it’s whimpering, it’s gnawing at its paw, it’s hiding, it does all the things that, very similar things that you would also do if you were in pain. Of course, you could have a toddler in pain. The toddler can’t tell you directly, but you know exactly when your toddler is in pain, right? So, there are many cases...Or a person had a stroke, who can’t talk, who is aphasic, again, you know when this person is in pain. Just because he can’t talk doesn’t mean that they don’t have conscious states. We know the nervous system is really very similar. If you look at a little cubic millimeter of brain tissue of a monkey, a dog, a mouse, a human, it would take an expert, it would take expert neuroanatomist to tell the difference, because if you
certainly look at the elementary units of computation, if you look at neurons, they are everywhere the same.

Yes, there is tiny differentiation depending on your ecological species, where you have evolved to, but those are really very, very tiny perturbations upon a common theme. We also know this because you can give your dog pain medication, right? So, I run a lot in the mountains and when I run more than two or three hours in the mornings, my dog, she loves to run with me, but, afterwards, she is in pain, and I give her aspirin, and the aspirin seems to help, just like it helps me. So, we know that the structures even down to the receptor level are either identical or very similar. So, for all of these reasons and, of course, for reasons of evolutionary continuity, it really would be mind-boggling if those guys, animals, didn’t have something that only we had. That would be mind-boggling, that would be...there really is no evidence for it. It’s really a distinctiveness in particular with respect to self-consciousness, the ability to reflect upon ourselves, to know who we are, to know that we are going to die; those things tend to be more unique to us, but not the basic level of consciousness. What is really scandalous is that we don’t know, so there are many things that we don’t know about consciousness.

One of the things we don’t know (is) how far consciousness extends across the animal kingdom. So, if you look at a bird, they can have very complicated behaviors, if you look at parrots, for example, Alexander, the famous Gray parrot of Irene Pepperberg who just died recently. They can, of course, talk to them to some extent, they can hide things, they can remember up to five thousand different caches where they put their food. If you look at tuna fish, they can have very complicated social organizations. I once struck...I come from an Institute where we study the visual systems of flies and bees. If you look at bees, it is amazing what they can do, so, for example, what you can train a bee to do, you can train it as quick as you can train a monkey, to fly, for example, inside a maze. At the entry of a maze, you have, let’s say, a red light.
The bee comes to a branch point in the maze. One branch has a green light the other one has a red light. The bees remember that at the entry was a red light and now at each branch point it has to take the exit that is marked with red. And you can cascade this, so you can have more branching points, every time you have a red and a green, and it always will take the red. Next try, you put green, and, now, it remembers: “I’ve got to take a green.” Then it generalizes, in a few trials, you use sugared water to teach the bee, it generalizes by, rather than putting a color at the entrance of this labyrinth, let’s say you put a horizontal or vertical grating, it knows now that, if at the entrance there is a vertical grating, it’s got to remember to take the exit marked with the vertical grating.

So, that’s really very, very complicated behavior. A bee only has eight hundred fifty thousand neurons—that’s not a lot—but we have no idea that these guys, that it doesn’t feel like something to be a bee. It’s not like in the movies with the insects; it’s not that they have voices inside their heads, but how do we know that these bees, that it doesn’t feel like something to be a bee. And so, right now, I assume that, unless there is evidence to the contrary, there is a reasonable chance that bees are conscious because of their complex behavior. The fact is that we really don’t know. We have this strong intuition of, “For heavens’ sake it’s just a bug,” right? But that is just a bias not really based on any solid evidence. What can we say for certain about consciousness? Well, we know for sure consciousness seems to be associated with some types of complicated systems, not all complicated biological systems.

Some of you may not know but down here in your gut there are roughly two hundred million neurons, down here, right, it’s called a second brain or the enteric nervous system. It controls things you would rather not know about, you know, peristalsis, taking all the stuff down, and the plumbing, but the fact is that these are bona fide, genuine neurons, they have neurotransmitters, they have electrical activity, they communicate, there is a network down there. They are very complex. There is not as many as up here, but two hundred million is a lot of neurons. It’s more neurons than,
for example, in a dog. But, as far as we can tell, the neurons down here, they aren’t conscious, or at least they aren’t telling us. Occasionally, you get vague feelings, you get feelings of nausea, etc., but we know those are actually communicated by a few couple of thousand neurons that make up the vagus nerve, and the sensation actually happens up here in the insula, in part of the cortex.

So, you have an example of a wonderful, complicated neural network that seems to do all of its business without giving rise to any conscious sensation. Your other complicated systems, for example, you’ve got your immune system, your immune system is very complicated, it works day-in, day-out, but it does its work totally in silence. You don’t have any sensation. Every day it fights off viruses and bacteria and other nasty stuff, but you don’t feel it. It is very complicated; it learns. You get one exposure to a virus, you get immunized for the rest of your life, your immune system will remember this. So, it’s as complicated as the nervous system, but, for some reason that we don’t know, it doesn’t give rise to conscious sensation. Only the neurons up here seem to do it. Why? We don’t know yet.

Consciousness doesn’t require behavior, so that’s to say...some of you may have seen the movie “The Butterfly and the Diving Bell.” So, there you saw the editor of “Elle,” this French fashion magazine, who had a stroke and then had what we call the “Locked-in Syndrome” where he could only move his eyes up and down. Or, even once you lose that...he was so conscious, he could even write a book about it, dictating it to a secretary by blinking his eyes. So, we know from many clinical cases that, certainly as an adult, once you are conscious like any of us, (if) we now have one of these nightmarish accidents or strokes, we are still conscious. So, we don’t need to actually act to really be conscious. We don’t seem to require emotions, certainly none of the strong emotions, like love and hate and pleasure and fear. That they are certainly necessary, they are essential for us to survive in the society, to survive danger, etc., but you can certainly seem to survive at least in a clinical context without them.
Again, we know this from patients, some patients that have lost part of their brain, for instance, with particular types of prefrontal lobe lesions, and then they are flat affect. So, they don’t seem to mind that half their body is blown away by a mine, let’s say something, they just don’t seem to mind. They don’t seem to have a lot of emotion. They’re clearly conscious. So, while emotions are necessary to survive for the long term, they don’t seem to be really required for consciousness per se. We don’t require language, we know that from children before they talk, we know this from people how have strokes, we know this from animals, nor do we require self-consciousness. Once I gave a talk a couple of years ago in France, and a lady came up afterwards and said, “Well, you know, that is all very nice Professor Koch, but I don’t think that you can ever convince me that any animal is conscious.” And I told her, “Well, that is very nice, but I don’t think you can ever convince me that you’re conscious.” And she was taken aback by that and then she realized, of course, that the only way that I know about her consciousness is also through behavior, one particular type of behavior, she talks. Well, that’s a behavior, just like grunting, or whining, or running away, those are all different behaviors. And so, yes it is true, historically, philosophers and linguists, of course, have given, the only behavior worthwhile that I am willing to consider is language, and, unless you talk, you are just not conscious. But, that’s of course a very narrow definition of consciousness and of what it is to be a sentient creature that happens, fortunately, to coincide only with humans.

And so, by that definition, of course only humans are conscious. We have this deep atavistic desire that we are always special, that we are always special and different from other creatures, but, certainly, if you look at the signs of it, the biology of it, there is no evidence of that. And we know from split brains studies done by Roger Sperry, Mike [Ghazni], and others, certainly, consciousness can occur in one brain hemisphere. So, whatever it is, it seems to be replicated in both halves, yes, there is specialization in general, the eloquent hemisphere is the left one, and the right one seems to be more specialized for vision and other things, but, by and large, both hemispheres are
conscious. Whatever the mechanism is, it is replicated. Very important, particularly for people with global theories of consciousness, we know that if you lose a specific part of your cortex or your cortical thalamic system, you get very specific deficits in consciousness, not a loss of consciousness, as I mentioned earlier when you have a lesion or a localized distraction in the brain stem. But, let’s say, if you lose part of, for example, your fusiform gyrus, you may well be like the title character in Oliver Sacks’s book, “The Man who Mistook His Wife for A Hat.” You may be unable to recognize a face as a face, or you may recognize it as a face, but you don’t recognize it as your wife’s face. Or you can see the face, you can see the nose and the mouth and the eyes and the ears, but you are unable to put it together into a conscious percept of a face. And so, that tells us that there is some very local aspect about consciousness. That’s a very important fact, because that tells us if you take locally take away this part of the brain, you have a specific deficit in consciousness.

Once again, this really strengthens the connection that there is between the brain and the mind. Also, if you’ve ever partaken in a clinical operation for some deep brain surgery or something else, then, again, you realize, very quickly you realize the intimate connection between the mind and the brain. We don’t know the final answer, but for sure we can say that the mind, the brain are deeply connected. You go inside the brain, you stimulate the brain, right there and then in the operating room, and the patient will suddenly say “oh, you just tickled me,” or the patient will, depending on where you are, will jerk the hand or do something. So, we know there is a very intimate connection between the two. What is it that we don’t know?

As I mentioned, we don’t really know the minimal brain size necessary for consciousness to occur. We don’t know whether it’s...for example, it would be great if it’s present in model organisms like C. elegans. The C. elegans, the round worm, the most popular biological model organism. Does it have consciousness? We don’t really know. Most biologists are willing to give consciousness to mammals because we are so
similar, but when you go away from animals that don’t even have cortices any more, it becomes much more difficult. What about Paramecium? What about Drosophila? Right now, we don’t know. We don’t know whether consciousness requires a body. This is particularly important if you consider about machine consciousness. We don’t know whether, well, it’s the same thing, whether consciousness just depends on the functional relationship. So, in AI, in Artificial Intelligence, there is a belief, for example, that in order to replicate intelligence, all you need to do is to replicate the functional relationship. It doesn’t really matter whether you use neurons or whether you use silicon or whether you use optics, as long as you replicate functional relationship, you can instantiate any information-processing machine. Many people believe that about consciousness, but we don’t really know, that’s just a belief at this point. But, if you can replicate all the functional relationships, let’s say, that you have in a human brain, that then you would also get consciousness in a machine. We don’t know, it’s reasonable, but consciousness is so strange, we don’t know whether that’s going to be true or not.

What we really lack right now, what we totally lack, is a theory of consciousness, although there are people like Julia [Ciccone] working on it, we don’t have a general accepted theory of consciousness that tells us from first principle, what is it that we want to know. And what is it that we want to know is, which systems under which conditions have this thing called consciousness. You know, I’d like to have an operational means, like some sort of conscious-o-meter that I can point at some of you in the audience, so I can tell whether you are asleep or whether there is any consciousness there whatsoever. I’d like to know in a baby, I’d like to know in a newborn, I’d like to know in a fetus, I’d like to know in a person, let’s say, that can’t talk, or in somebody like Terri Schiavo. I’d like to have a practical way of measuring it. I’d like to know in a dog, in a cat, in a mouse, in a squid, in a fly, in C. elegans, I’d like to know it for a Macintosh computer, for a PC—probably feels horrible to be a PC—I’d like to know it for the internet, for example, is the internet as a whole conscious?
Right now we have no way of even thinking about those things, but we want to think about them, because we need to understand if science is supposed to have a complete description of the universe, and we must include the fundamental aspect of the universe, mainly that it contains a conscious observer, and we have to see our consciousness fit into everything else, and so we want a theory about that. Alright, so what Crick and I, Francis Crick and I, advocated that it’s nice to talk about all these general things - but, of course, science is all about the details - so, for now, let’s go away from these big questions and focus on something that we can do today, namely on focusing on the neuronal correlates of consciousness. So, they are now called the NCCs, the neuronal correlates of consciousness are the minimal mechanisms in your brain that are necessary for any one conscious sensation, let’s say yellow. What is it that you need in your brain in order to see those yellow squares?

For instance, of the twenty billion neurons in your cortex, there are roughly fifty billion neurons in the cerebellum, this little brain here at the end, at the back. You need your cerebellum—fifty billion neurons, that’s a lot of neurons—you need your cerebellum in order to see yellow. There is no evidence for it. In fact, there are sorts of evidence, you take it away, the person can’t play piano or other things, but there really is no difference to their conscious sensation. You need the eyes, you need your eyes to see yellow squares. Well, to see and hear you do. But, of course, tonight you may dream of yellow squares, and when you dream, of course, your eyes are closed, and, in fact, all your input is pretty much shut down, but you have very vivid dreams. Tonight, I had some very vivid dreams, and you have those dreams in the absence of retinal input, and all of our dreams tend to be visual, unless you have a stroke in your prime visual cortex, they tend to be visual. So, clearly, yes for normal forms of things you need your eyes, but to seem conscious at all, the eyes don’t seem to be essential. And so you can ask for different parts of the brain, so, for example, Professor Crick and I proposed in this “Nature” paper ten years ago that you don’t need your primary visual cortex, that your prime visual cortex, also known as V1, is essential for many forms of seeing, like seeing
right now, but that's not where consciousness actually is generated, that's not where visual conscious percepts are actually generated.

So, in principle, you can ask that, not only at the level of area, but at the level of neurons. So, let's see, if you think that it's a high-level visual area, talking to a prefrontal area, and this area needs to communicate, you can ask, “Well, do you need all the neurons in that area?” Probably not. There are a hundred thousand years neurons per cubic millimeter, there are probably at least in each given cortical area a hundred different cell types. We can now begin to identify them. You probably don’t all need, you probably may only need the large projection neurons, of which there are many types, and maybe you only need the subtypes of those. So, you really want to chase down consciousness down into the circuits, you want to track down to the level of the circuit, you want to understand which synapses are necessary, which molecules are necessary, because then you can interfere, you can deliberately turn things on and off, and that’s a very powerful technique that we can begin to exploit. So, focusing on the NCCs has its advantage—you don’t worry so much about the hard problem in chemscience. You don’t have to explain it, you just try to chase down the [collage of conners] in a particular brain of a particular individual, human or an animal. So, we know, and there is a hundred years of psychology behind this, of course, most famously associated with Freud - although he stressed the sex part a little bit, too much for most of our tastes - is what Francis and I call these “zombie” agents in your head. Most of the stuff that happens, and you realize this when you get older, the large part of your life, the patterns of your life, that start off with getting (up), hitting the alarm, getting out, showering, tying your shoelaces, driving, interesting things like dancing and climbing and playing tennis, all those rapid-sensory motor activities we do without consciousness. Consciousness, if at all, happens afterwards.

For example, we know this extremely well for eye movement. It is very well studied, and, for most eye movements, you move your eyes as roughly as often as your
heart beats, a hundred thousand times a day. Most of the time, you have no idea you are moving your eyes, in fact, you can show this with very nice experiments. You can manipulate your eyes much more precise than you can think, than you can see. In other words, your eyes can correct for things that you don’t even consciously see. So, there is a lot of evidence that your eyes do whatever they do all the time without consciousness being involved at all. This turns out to be true for most things, for example, you don’t really know how you understand what I say, assuming that you understand what I say. I talk in this funny, I talk like my Governor, actually, right? I talk in this funny accent, and you hear them, and, somehow, what I say makes sense, but you don’t access the language-processing module. Same thing when I talk; it’s not that I am trying to say something in my head, let’s say in German, and, then, I put a verb and a subject and I put them together grammatically correct, and then I send them off to my larynx, no. I have a vague idea what I want to say and, then, next thing I hear are these words coming out of my mouth. Yet, I do it all the time; we do it effortlessly. We reach out and grab things, and we know from robotics how difficult it is for you to reach out and grab it and do weird things with it; and I can grab a weight or I can grab a feather or an egg, and I can do it all the time effortlessly, and you can show this in patients. The patients are unable, for example, to recognize things or unable to see things, yet they can still reach out and grab.

What that tells us is that these patients, for example, one famous patient I’m thinking of, she had a carbon monoxide poisoning, so she was unable to recognize it as a bottle. She couldn’t see whether the bottle was horizontal or vertical. She couldn’t see whether it was big or small, yet, she could perfectly well reach out and grab it, and either she would do this or she would do this. Otherwise, she would shape her hand accordingly to the shape of the thing she had to pick up, yet she didn’t have access to that information consciously, because the part of her brain that mediates conscious vision was destroyed by the carbon monoxide poisoning, but the carbon monoxide poisoning, fortunately, left the other part of the brain intact, namely the part of the
brain that’s necessary to mediate these unconscious sensory motor actions that we need for doing, basically, everything. There is one very interesting patient, described in a wonderful book, who lost—it’s a terrifying story because it could happen to all of us—who lost, at age twenty-one, he got a flu, he got a virus infection, he lost all of his deep, sensory neurons. So he couldn’t sense his body anymore; he didn’t know if his hand was like this, like this, or like this. He lost all of them, permanently damaged, gone, kaput. First three months he couldn’t move at all anymore. He totally collapsed, he was on his bed, he was basically paralyzed, but, in fact, he wasn’t paralyzed. His motor output was perfectly well working; it was just that he couldn’t sense where his body was. And then, for the next two years, he was very, very determined to learn how to walk by looking, to look at his feet, and he had to position them. He had no idea where his foot was positioned in space. So, he had to learn to move by looking at his muscle and doing everything consciously. Folks, he talks about when there was a power failure, the lights went out, he immediately collapsed because he didn’t see his body anymore. He had no way of knowing whether his body was like this or like this. That tells you that if you only had consciousness, we would be a wreck, we couldn’t do anything, we couldn’t get out of bed, literally.

Alright, so it is very interesting if you only think about the neuronal corollary of that; where is the difference in their brain? Because we know from single-cell recording in monkeys, and from fMRI, from EEG, that all those, for example, patterning in moving involves neurons in our cortex. So, the fact is that the nerve cells in our cortex proper that are active, that by itself is not sufficient by itself to give rise to a conscious sensation; it has to be something above and beyond. It can’t just be, as people used to think, OK, anything that gives rise to electrical activity in the cortex is consciousness. That is not true. It has to be a particular type of electrical activity. Let me show you two or three nice experiments. This is one of the oldest techniques in psychology. What should you do is just look at it, don’t move your eyes, keep them open and look at the colors here. Just as I talk, don’t move your eyes, keep looking at the colors, and what do
you see now? Ok, let’s do that again. You should tell me what you see. Once again, just keep your eyes, it works best when you keep your eyes totally fixed. It’s called an after-effect. It’s a color-after image. And it’s really quite vivid if you go back and forth, it’s really remarkable, right? Now you see colors here, but, of course, the colors are not really there; it’s all in your head, not in the world. And they slowly fade. By the way, this illusion contradicts the naïve realist, because the naïve realist assumes, clearly, there is a world outside there, and the world is mapped onto my brain, and that’s all there is to it, but this tell us that things are a lot more complicated. It depends on the history of the world, because if you now look at it, it is perfectly gray, but a minute ago, after you looked at the color image, there were colors in there, and you can measure these colors. These colors are real to you, and you can put people in magnetic scanners and see that these colors actually exist in your head, not in the world, but you see them, of course. Remember this illusion, again? While I talk, do the same thing, keep looking down here at the “B,” because we can now ask the question, which seems like a weird question, but you can now ask it, do you need to see the yellow squares in order to get an after-image to the yellow? So, here, you should be able to see an after-image, right? You see the bluish squares? Very nice, right? By the way, there are certain things that we don’t understand and they wax and wane. They are strong, then weak, and then strong again; we have no idea why that is the case. You can ask the question, “Do you need to see the yellow squares in order to see the after-effect?”

You can do this experiment, this is a known experiment, you can do very simple, essentially, what you can do, you can go back to the illusion, so, let’s say, this disappears for five seconds, so I ask you to monitor—I give you two buttons, whenever one yellow disappears, you push one, when the other yellow square disappears, you push the other one—and, then, you release it when it reappears. I can try to do a comparison. Let’s say this yellow disappeared for five seconds and this yellow is present, and, now, I can immediately remove the image and see the after-effect. Does the after effect on this side, is it shorter and less intense than the after-effect here? The clear answer is no. So,
the clear answer, you do not need to see consciously the yellow square in order to have an after-effect; it doesn’t depend on it. That is really interesting, because it tells us something about the brain, the organization of the brain. By and large, these after-effects are phenomena of the retina, and so this tells us something we knew already, in this case, that you don’t see with your eyes. You need your eyes in order to capture the photo and convert them into electrical activity, but, ultimately, consciousness happens in high-level cortex, and your retina captures the image. The intensity of the after-effect just depends on how long the image was on the retina. It doesn’t at all depend on something up here.

So, it tells us, again, consciousness is not this global phenomenon that mystic and holistic people talk about, it has to do with specific parts of the brain. Now, you can do more sophisticated after-effects. You can do (this is dated, this illusion, I’m sorry, I have to update it), who do you see here? You can see why I need to update it. Who do you see here? C’mon, c’mon guys. How many people see Kerry? And how many people see Bush? How many people see Kerry here? Anybody? And how many people see Kerry here? Ah nice, many more, very nice, alright. Alright, so what happened here, I showed you three images, but actually there were only two different ones. This was fifty percent Bush and fifty percent Kerry; it was morphed. So, now, of course, this works best, everybody sees this slightly different, so I really have to do this for each person by himself, everybody has to morph, because you have different points where you see morph, where you see it now fifty percent Bush and fifty percent Kerry. You know, it’s for a public demo. So this was fifty-fifty, this was pure Bush, a hundred percent Bush, alright, not for much longer, and this again, here we went back to the first image—this again, is fifty percent Bush, fifty percent Kerry—however, now, if I asked you, may more people raised their hand that they saw Kerry than before. First, I showed you fifty percent Kerry, fifty percent Bush, then I showed you a hundred percent Bush, then I showed you again fifty percent Kerry, fifty percent Bush, but many more of you saw Kerry than Bush. What happened? You adapted. Because I showed you the hundred
percent Bush images; we don’t understand the mechanisms of faces yet in the brain, but we adapted to it, so now with the same input you are much more likely to see Bush than Kerry, so, tonight, it is called the face-identity specific after-effect.

And what you can show using [INRS] technique, we published in “Neuron” a couple of years back, you can show that for this illusion, in order to get the face-specific after-effect, you do need to see the face. In order to see the yellow, you don’t need to see it, in order to get the after-effect, for the face, you do need to see it in order to get the face after-effect. That tells us that the place in the brain where the conscious perception for faces happens in an area, we think, called, or what Nancy [Kammerer] calls confusion from face area, is also close to the area where the after-effect happens, which is very different than the after-effect for color; that happens much earlier. So we can do what [Bueler Miller] calls psychoanatomy; without opening the brain, we can locate where the different parts of the brain that generate the after-effect and that generate consciousness.

Ok, let me show you another experiment. For this experiment, you have to be very quiet because you have to count, and it is going to be difficult for some of you, so, you should really concentrate and don’t talk to each other and just do this task. What I am going to show you is a video; some of you may have seen it before. This is a video to show the task that you have to do. There are two teams; there are three people in white t-shirts and three people in black t-shirts, and you have to track, they each have a basketball, and you have to track the people in the white t-shirts, and how often they pass the ball to each other, alright? So just be quiet while it happens and concentrate on the task. How often do the people, the three people with the white t-shirts, how often do they pass the basketball to each other? Alright, so how many people saw the grim reaper? You’re too alert. Did everybody see the grim reaper? Alright, for those of you who didn’t, let me show you again. The point here is to show the relationship between attention and consciousness. Usually, what you are attend to you are conscious of, and
if you don’t attend to something, you may totally miss it, although it is right in front of your eyes; this is, of course, a trick magicians use everyday. This is the same video. This was done by Dan Simmons and this was refilmed by PBS. So there you can see the grim reaper. Some of you, I know, missed it. The first time you see it, many, many people will miss it. This should be very disconcerting, because this figure is present three, four, five, six seconds on the screen.

But, of course, I misdirected you; I told you to look at the people in white, to concentrate on this difficult task; this is just to make the point that if you don’t attend to something, you may not see it, and, this probably now explains the significant number of accidents, for example. Accidents happen in full daylight, no foul play, no alcohol, no drugs, nothing. People just run into things. By the way, I’m serious now, this form of blindness, or related form of blindness, called “inattention blindness,” happens two or three times more likely if you are on the cell phone. Because on the cell phone, you are totally engaged in this conversation, particularly if it is emotionally engaging, and you may totally miss things right in front of your eyes because your attention is elsewhere. From a psych point of view, you should definitely not be using your cell phone while you drive.

However, although traditionally attention and consciousness have been conflated in most theoreticians who have thought about it, think they are the same, I think that there are two different processes with different functions and different implementations in the brain. This shows that we begin to understand attention pretty well, so, I’m an engineer, so I’ve worked with people, with my old advisor, Tony Poggio, and here we have, an attentional system this is meant for computer vision, just to show you what we think the function of attention is, so it’s a video outside of MIT, outside the I-square, the AI lab, you can see this is the video, and this is the intentional search-light; it’s called the [same search attention], so it analyzes anything in the search-light, and then it outputs over here what it thinks it is, a car, it gives you the confidence that it has,
the confidence it has in its classification. It does a difficult job known as object categorization, which is more difficult than object identification. And also, the function of the...this is just in the machine vision context, a biological-based model of machine vision, the function of attention is to select - and we all have it, flies have it, you can show very nice experiments in cats and dogs, and monkeys have it - so the function of attention is to select this huge amount from the flood of incoming information, to select the subset and to make that subset accessible for higher-level processing. Usually, this coincides with attention, but it doesn’t have to, with consciousness, but it doesn’t have to.

And now, a whole set of experiments where you can show, you can disassociate attention-consciousness. One in particular is really very sexy. This is a technique a student of mine invented called the continuous flash suppression where we can do things just like a magician can do, but we can do them reliably, reproducibly in a magnet. So what happens here in this particular technique...let’s say, on your left eye, you’re seeing this angry face, which is a potent stimulus. In your right eye you see these flashes, they’re updated every hundred milliseconds, of, in this case, called mondrians. These are just abstract random patterns. And what you’ll see, and we do this now for three minutes typically, and people—almost everybody—sees only this. So, again, we can monitor, we ask you two buttons and press the left button if you see a face or a higher valency of the face and the right button if you see this continuous flash suppression. If you keep both eyes open, you’ll only see this. This is a very vivid stimulus; it grabs your attention and changes. This one is stationary. If you blink with your right eye, you immediately see the face, that’s not the issue. It’s just when both eyes are present, the right eye input powerfully suppresses the left eye. It’s called continuous flash suppression. Now you can use this technique to do really interesting stuff.

This one is the most interesting one because it pertains to sex, something we all care about. So, here, what this has done, showing down here out of the Minnesota
group of Jiang He. So what he does is shows you pictures of naked men and naked women, let’s say on the left, so let’s say this view on the back for modesty, it’s covered up here. So, let’s say this is the left eye and this is what the right eye sees. So, let’s say, the left eye sees on its left side a picture of a naked woman, on the right eye, you have the same picture, but you cut it up in little squares for control. In the right eye, you see these flashing, these random-colored squares. And now your mission is to say, for example, was the naked woman on the left or the right side? Or you can do a different control experiment, was it a man or a woman, for instance? And, here, people are chance, [deep prime] zero; they don’t know, because it’s a powerful math, they don’t know whether, for example, the naked man or the woman was on the left or the right side. They don’t have access, they don’t see anything, because the only thing you see are these colored squares. Yet you can show (ok, I missed a slide), you can show that by doing a little attentional measurement, that, in fact, people attend to this side where the naked person was. And more interesting, even, you can show that this attention depends on your gender and your sexual identity. So, if you are a straight guy, you only pay attention to this image if it is a naked woman; if it’s a naked man, in fact, you seem to avoid it, interestingly. If you are a gay man or you are a straight woman, you tend to pay attention to the image of the naked man. But the disconcerting thing is, is you don’t know this, you don’t see it, because, consciously, if I ask you, “What you see?” all you see are these colored squares, because this thing on the left is hidden under these colored squares. It’s really interesting because it gives us access, it’s a little bit Freudian if you want, it gives us access to unconscious processing, in this case, it relates to your gender and to your sexual identity, which is pretty cool.

Now, you can do other experiments with this technique, so, for example, this is an experiment out of the London Group of [Lavie and Rees], you fix it here in the center, and, let’s say, in the left eye you put this images of a tool, you know, some tools here, and in the right eye you put these flashing squares again. And, once again, using fMRI, now, you can peer inside the brain, and you can see in the back. (So you can feel there’s
a bump at your back here, you should feel it; the back of your head there is a bump. If someone hits you, whacks you there with a baseball bat, you’re going to get what cartoon characters get when you hit them over the head. You see stars, because you are directly stimulating your visual cortex.) There are different areas here called V1 and V2, V3, etc. Anyhow, so, what you can show at this invisible stimulus attacks attention, these invisible tools, although, once again, people had a chance to say, “Are they here or are they here?” They don’t have access to this information; something in the brain does and pays attention to it. So what this really quite nicely shows, and I think most experts now are in agreement with this, that you can pay attention to things without you being conscious of them.

So that means, just like in any science, when early on you conflate terms, and later on as the science becomes mature, you realize that they are actually different things. And consciousness and attention are different things that require different explanations and different neuronal mechanisms. Let me finish and come to two neuronal mechanisms stories. So, one is, ultimately, we need to understand this at the level of neurons. Almost everything we know about neurons is from animal experiments, but there are very rare conditions where you can study nerves in humans. We do this since many years with a wonderful collaboration with the neurosurgeons and neuroscientists who are based in Israel and at UCLA, Itzhak Fried. The context is...you know what those are? Those are human neurons talking to each other. They sound just like the way neurons do in animals. What this is, is recorded and amplified and differentiated pulses from individual nerve cells in a part of the brain called the medial temporal lobe, which includes the hippocampus and the amygdala. The context here is that people with epileptic seizures, some of them the drugs don’t work any more, and then what the neurosurgeon has to do is to go inside the brain and scoop out of the part of the brain that gives rise to the focus. It is very successful. It’s invasive, but it is a successful therapy. In many of these cases, you don’t know from the outside where exactly the seizure originates. So, then, what the surgeon has to do is insert these
electrodes. Here you can see inside the hippocampus, here, twelve of these big electrodes inside the brain.

What Itzhak Fried does, since many years now, is he hollows out this electrode, and then in here he inserts these microwires, nine microwires, we measure with respect to one of them. So, essentially, we have eight wires per microelectrode, and there are roughly ten microelectrodes, so we have, on the whole, a hundred wires inside the patient’s head, maybe for three, four, fix, six days, seven days (so, here you see a patient) until the patient has sufficient seizures so the neuroradiologist can pinpoint, can triangulate and pinpoint the seizure. Then the electrodes are taken out. With their permission, this allows us now to record from nerve cells, because, ultimately, ground truth for the brain is neurons; it’s not imaging. That’s cool, but, it is very cool, but we really need to understand it at the level of individual neurons. So, here you have a patient, and he is looking at images on the laptop, and he can tell you what he is seeing, and we can record from his nerve cells...old, the very ancient bacterium rhodopsin and you have shades of it in your cone, in your rods for the receptors in your eye. So, essentially, what you can do, you can essentially insert a photoreceptor that has maximum sensitivity in the blue, four hundred fifty nanometers or so, and you insert that now using a virus vector. So, you insert the gene for blue photoreceptor inside, let’s say, a retrovirus, which is like an attenuated HIV virus. You get that virus inside the brain, for example, by injecting it (it’s one technique) and now you add a promoter to this virus where that virus and this gene only gets expressed in the particular type of neuron, for example, layer 6 [pranal] cells.

Or, in these experiments, in the lateral hypothalamus of mice, this is a part of the brain that we thinks is involved in us in sleep regulation, sleep and wake transitioning. It’s part of the job of the hypothalamus. In particular, there is a neurotransmitter that now we know in animal studies and human studies involved called orexin, hypocretin. So what this group of Carl [Desaroff] at Stanford did, they took a virus, they inserted this
photoreceptor and they inserted the promoter for the orexin neurotransmitter and they injected it in the hypothalamus. Now, this virus only affects neurons that have the orexin, and in those neurons, it now expresses as a blue photoreceptor. So, now, deep inside the brain, you’ve got neurons that have a photoreceptor. That is pretty stupid, because, of course, there is no light there, unless you insert optical fibers and you bring the light to it. Specifically, you bring blue light to it, because blue light, just like the photoreceptor here, the neuron gets excited, the neuron fires, so, now, you can now artificially, deep inside the brain, using pulse of light, which, of course, you can time very exquisitely, you can directly time, and you can get the animal to wake up. So, that is what showed in this “Nature” paper that came out a couple of month ago, late last year.

We can wake the mouse up either from REM sleep or from non-REM sleep by giving pulses of light, and the beauty of it - they have another gene they modulated - halorhodopsin, it’s essentially a chloride pump, you can inhibit the neuron. So, now you can play the light organ, you can play the light on fantastic. You give blue pulses of light, you turn the neurons on, you give yellow pulses of light, you turn them off. You can do it much more specifically than by just inserting an electrode in the brain, because now you are using the genetic regulation, so you can combine optics with genetics to really specifically target a specific group of neurons. It’s really cool. So what is the long-term strategy? So, we need to develop robust animal models of consciousness, in monkeys, in particular monkeys, in mice, but maybe also in other animals like flies. We, ultimately, we need to identify the groups of neurons - I call them coalition -the groups of neurons that give rise to specific conscious sensation like where is consciousness for blue or for yellow or for pain? And, then, ultimately, what we need is...we want to interfere with them because we want to move from correlation to causation. fMRI right now is correlation; you see, every time I think of faces, this part of the brain lights up. I like to move to causation, as any strong theory should. And then, ultimately, we need a fundamental scientific theory of consciousness.
Lastly, this research is already beginning now; general science research is only beginning now—it should be more in the future—have implications for the way we live and the way we comport ourselves. Once we really understand the science of consciousness, once we understand the mind-brain problem, there are a certain number of very tough problems that have arisen, or that will arise through advances in technology, that we have to ask and answer: for example, who can live, what about an encephalic infant, what about an infant where you are sure this infant, based on our best current scientific knowledge, will never have any conscious sensation? The infant may live, may live maybe even a full term, but will never have a conscious sensation. Is it ethical under these conditions if that person can never be a person to use the organ from this infant to save another infant’s life? At the other end of life who can die? I mentioned already Terri Schiavo was terribly controversially politically or religiously; it wasn’t really controversial scientifically, and, as I mentioned, there are roughly fifty thousand of those people around today. What do we do with them? What should we do with them? Since I do this research on consciousness, I turned into a vegetarian since I don’t feel I myself can justify raising conscious creatures, like lambs and pigs and other mammals, under pretty atrocious conditions, to eat their flesh. Is this something more of us should do? Or should we build, if we really have a theory of consciousness, which will enable us, ultimately, to build conscious machines, should we build conscious machines? Because there are all sorts of consequences that entails, so I hope to leave you with those questions, and, just with the fact that we can now study these age-old questions of philosophy; we can now study them in a scientific context. Thank you very much for your attention.