

Key:

I = Interviewer

R = Respondent

Unclear: [unclear + timestamp]

Talking over each other: [over talking]

Welcome to IFLScience – The Big Questions. The podcast where we invite the experts to explore the biggest mysteries of science with your host, Dr Alfredo Carpineti.

I: Welcome to IFLScience – The Big Questions, a series where we ask experts some of the most pressing mysteries of science, technology and humanity. I’m your host, Dr Alfredo Carpineti, IFLScience’s Senior Science Writer. It is wonderful to welcome Professor Liad Mudrik from The School of Psychological Science and Civil School of Neuroscience at Tel Aviv University. The question this time is, what is consciousness, and how does it arise? Professor Mudrik, it’s great to have you here. Could you tell us a little bit about yourself and your work?

R: Sure. First of all, thank you for inviting me. Happy to be here. I work, as you said, at Tel Aviv University. My lab, which has wonderful students, that’s a great privilege, works on consciousness. Trying to understand first how it comes about, so how it emerges from neural activity. Second, what are its functions, so for what types of processes do we need conscious processing or what type of processes could also take place without us even being aware that we are doing them. We are also interested in that respect in consciousness’ role in decision-making and their voluntary action. Finally, we also ask how does our knowledge and beliefs, how do they affect our perception? How do they affect our conscious perception of the world around us?

I: Thank you very much. That is very, very interesting work. Let’s start with The Big Question. What is consciousness? How do we define it?

R: That’s a huge question. Philosophers and scientists have been struggling with it for years, centuries, trying to understand what is the intricate phenomenon that is so intriguing and so interesting that on the one hand, and this is a huge paradox, there is nothing that is more immediate to us. This is something that we don’t need any mediators in order to experience. We just have it. From the moment you open your eyes in the morning until you go back to sleep, and when you have dreams at night, you are conscious, and no one should explain to you what your consciousness is. You just have it. But, on the other hand, this is probably one of the least understood phenomena. Although it is so close to us and so intimate, we have such intimate knowledge about it, we don’t understand exactly how to define it. We don’t understand exactly how it comes about and what are its neural correlates or neural mechanisms that give rise to it.

That was just kind of a disclaimer to say that whatever definition I’m going to give you will be something that people debate over. This is something that people have been trying to understand for a long time. When I talk about consciousness, I am not talking about self-

awareness or political consciousness or anything of that sort that is very high level. Our strategy, and I think that many people in the field share that strategy, is to talk about consciousness in the sense of our experience itself, the way we experience the world. If you want, I guess the most well-known example is my favorite example, because I love chocolate dearly, is what happens when you put a chocolate cube in your mouth or whatever. The receptors on your tongue can process the information about that chocolate that you are just eating. They can offer examples or they can send information about the high level of glucose that you have in that chocolate. But on top of that, you are also experiencing the exquisite taste of chocolate. The question is what allows us as opposed to machines not only to process information but also to know how it feels to eat chocolate, to taste chocolate, or to listen to music. To see the redness or red or the blueness of sky and so on and so forth. This is the greatest mystery.

This phenomenal aspect of experience, what philosophers call qualia, are the... how it is like to eat chocolate, to smell a rose, to see, as they say, the sky at sunset. This is what we are trying to pinpoint. As you can imagine, it's not an easy task because you cannot just measure it. You don't have a ruler to say I have 30% of awareness now or I have this type. There are many, of course, empirical challenges, but I think that although this is probably one of the most challenging fields to study, it's very, very interesting and also very important due to the disorders of consciousness that are bound, and maybe the study of consciousness will help us deal with those. Also, of course, there are ethical and philosophical questions that accompany this type of...but I guess we'll get to that later on in the conversation.

I: Thank you very much for that. Something for me, with very much a background of hard science or physics. Actually, hard science. For me, it's just how do we ground it into reality? How does consciousness relate to our brain? Is it in our neurons? Is it the electrical signals among them? I understand that maybe [the answers] for even most of these questions are just we don't know yet.

R: We know some. We have some understanding. Let me first maybe say how can we study it even. How do we gain this type of knowledge because, as you say, as opposed to physics the phenomena itself is not easy to study. First, I must acknowledge all the people that came before me and my generation. The kind of founding fathers and mothers of the field who developed very elegant methodologies in order to create an experimental condition where I present you the very same stimulus, let's say a face, and I do it in a way that sometimes you will see it and sometimes you don't. Although the physical input remains constant, your conscious experience changes. We have many ways to do that. For example, I can flash a stimulus very quickly in front of you, or I can present a different image to each eye, which is a very unique experience because then your consciousness just switches between these two images instead of seeing them on top of each other. Although on-screen the very same information is being presented, your consciousness changes. The content of your consciousness changes. You have all these very interesting methodologies that you can use in order to create and an experimental condition in which we hope all other factors are kept constant, but the change is in your own experience. That is one way to study consciousness.

Another way to study consciousness is to look at states of consciousness. For example, you could ask what happens when I am having a dream versus when I am in non-REM sleep stage. Or I can what happens for patients in a coma or a vegetative state or locked-in syndrome or other types of disorders of consciousness as opposed to healthy wakefulness. All these types of contrasts allow you to ask what is the neural difference between these different states. Of course, another challenge when you study consciousness is how you measure it. As opposed to something you can simply observe, when we talk about consciousness, we are bound to rely on the report of the subject or some behavioral measures that tell us what he or she feels or experiences. We could also try to develop neural indexes for your experience, but they will, of course, always be validated through report. At the end of the day, I have no way of knowing what you experience other than asking you. Here we also develop different measures in order to make sure that the reports subjects give us are valued and are trustworthy, because these are kind of the building blocks of the experiments that we are using.

Today there are also new experiments where they try to track what subjects experience, for example, using their eye movements, so without asking them or inferring from their behavior in one way or another. This is how we study consciousness. Your question, I believe, was what do we know about the brain and consciousness. It's actually very interesting. In the late-nineties, Christof Koch, whom I had the privilege to do my postdoc with, and Francis Crick, the Nobel laureate, published a series of studies that in a way paved the way for the Neuroscience of Consciousness. I should say there have been psychologists and our scientists that worked on this project way before that, but many people regard that as a milestone in the history of the field because it kind of made it legitimate to work on consciousness as a key topic of research for neuroscience. Their suggested course of action was let's try to track the neural correlates of consciousness. The minimal set of neural activations that accompany any conscious experience, whatever that may be.

Then you can see in the history of the field many studies trying to do that using the paradigms I mentioned before. Some candidates indeed have emerged, so people have spoken about synchronization between different brain areas. A major candidate was frontal activation, so whenever there is activity in frontal areas or I would say the other way around, whenever subjects report consciously receiving something, we typically see frontal activations. The ventral stream in the temporal lobe, which is a group of high-level visual areas, has also been implicated in conscious perception. But with time, there have also been studies that try to challenge these findings and say maybe these are not the correlates of consciousness per se, but rather of the processes that accompany a conscious experience. For example, maybe it's a correlate of the report rather than of the experience itself. Currently, there are several series of consciousness each trying to suggest a different mechanism that gives rise to conscious experience. If you want, I'm happy to explain about some of these. All of them basically try to explain what in the brain allows us to experience. There are different accounts and, again, up to you if you want me to go into details about several of those.

I: Yeah, that would be interesting, just the vagary of how consciousness arises. I'm sure there is still a lot of work to do to pinpoint.

R: I'm also happy to tell you about one specific project where we are trying to arbitrate between some of these spheres.

I: Sounds cool.

R: Yeah. Let me first try to maybe describe four major theories in the field. I should say there are plenty of theories of consciousness. It seems like many people have very interesting ideas about how consciousness arises from neural activity. I'm here focusing on the four that are most, I would say, prevalent and cited, but there are many others. I won't do justice to the field by only focusing on those, but let's use that as a start. One of the leading theories is global neuronal workspace. It was originally suggested as a psychological theory by Bernard Baars and then developed into a neuroscientific one by Stanislas Dehaene and Lionel Naccache. The main idea there is that we have a central executive system which is called the global neuronal workspace. Its neurons are specialized neurons that are located mostly in frontal parietal areas. They kind of get activations from many processing modules that are spread throughout the brain. The idea is that unconscious processing unfolds automatically in these processing modules that are encapsulated.

For example, the area called the fusiform face area, the name is not very important for us, but the specific area that knows how to decode faces or process faces can do that unconsciously as well. When the information passes or crosses the threshold, then it enters into the global workspace that broadcasts it throughout the brain. That allows integrity of processing, it allows accessibility of the information, and this is when it becomes conscious. Whenever some kind of message or signal or presentation is strong enough to cross a threshold to be amplified by attention, it can be broadcasted throughout the brain, and that is the founding principle of consciousness for that theory.

Another theory called the integrated information theory developed by Giulio Tononi mostly, recently Christof Koch has also joined him, this theory is very interesting. It attracts a lot of attention but also a lot of criticism. I think one of the reasons is that its starting point is very different from other theories. It doesn't start with the data, it actually starts with phenomenology. What Tononi and his group have been doing is to say what are the necessary and sufficient characteristics of a conscious experience. They list from phenomenology, from introspection, they list five axioms which for them are the necessary components of any conscious experience. It has to exist. It is composed of different aspects. It is nevertheless unified, so it is one, it is integrated, it's informative. It also excludes parts of information that are not included in it. It's very definite. It is what it is as the theory goes. From the five axioms they derive postulates. Kind of conditions for what should be the physical substrate that could give rise to such an experience that has these five axioms. They try to formulate it mathematically and even developed a measure called phi that is supposed to quantify the consciousness of the system based on its structure.

It's a very complicated theory. I won't go into the details, but the idea is that based on the state of the system and the structure of it, you can derive an unfolded structure that takes into account the past states and the future states of the system, and that is the conscious

experience. It's an intriguing theory. As I said, it gains a lot of attention but also a lot of criticism. It's a very interesting option.

The next theory in line is the higher-order-thought theory. Interestingly, that also comes from philosophy originally. David Rosenthal is the philosopher who suggested this. With Hakwan Lau, a neuroscientist, this became a scientific field. The idea there is for a piece of information to become conscious. It's not enough to have the sensory areas decoded or represented, you should have a higher-order state that kind of points at it and tells you that you are experiencing this information now or this representation now. Without this higher-order state, you would process the information but you will not be conscious of it. You wouldn't experience it. The assumption is that this higher-order state that is related to metacognition will be located in frontal areas, specific frontal areas that are highlighted by the theory.

Last but not least is the recurrent processing theory, which is kind of the mirror image of the higher-order state one. It's basically a theory that says what is needed is the activation in the sensory areas itself. We don't need a higher-order state to point at that, but what is needed is recurrent processing to these states. We need some kind of a feedback loop that would make this specific that would allow some figure-ground separation and integration processes and would make that experience conscious. You see you have these different accounts and different theories.

One question you could ask is how do you differentiate between them and how do you arbitrate between them. In our work that we have done that is not published yet, but we're hoping to submit it soon, we also looked at how the field has been testing these theories. What we've seen is that each theory basically develops independently in parallel without much crossed off between the theories. Each theory proves itself in a way. We don't see enough challenges from one theory to another theory.

A very interesting initiative by Dawid Potgieter from Templeton World Charity Foundation, he once said let's try to use adversarial collaboration in order to test these theories. I had the immense pleasure of attending a meeting in the Allen Institute for Brain Research in Seattle led by Christof Koch, where he invited representatives of each of these theories. Then it was three theories, GNW, global neuronal workspace, IIT, integrated information theory, and higher-order-thought theory, to sit together with philosophers, with impartial scientists like myself. I don't belong to any of these theories. He said, let's try together to think of experiments that could test the theories. This started a very long process, and we are now halfway through after more than three years of working on this. We developed experiments that would test the global neuronal workspace and the integrated information theory with the people who proposed the theories.

I now, by the way, remember that when I presented their recurrent processing theory, I failed to acknowledge Victor Lamme, who is one of the major proponents of that theory. There it was Giulia Tononi and Stanislas Dehaene, two of the leading scientists in the field, who have not really engaged with one another for a long time, now sitting together trying to think how can we test our theories. From that came two experiments that we are now running in six different labs around the world. fMRI, NED-DG, and ECoG, so three different techniques for

neuroscience with a database that is going to be very large. About 500 datasets from 250 subjects that are going to be run on these experiments. The nice thing is that it's fully open science. All the predictions and the analysis plans are pre-registered and will become open to the public within two weeks or so. The proponents of the theories actually say we kind of confirm that the theory, that the experiments are well designed to test our theories. Here are our predictions, now let's see what happens. I think it took a lot of courage from these two people, Tononi and Dehaene, to put their theories on the line. They put their necks on the line, if you want, after 20 years of working on these theories or even more and saying this is what should happen. If it won't happen, at the very least I would have to revise my theory. I'm not deluding myself to think that after my experiments one of the theories will just say, okay, I'm wrong. But at least some aspects of the theories would have to be revised.

I think this is the way to make progress. I must here acknowledge the two people that work with me to lead this consortium, Michael Pitts and Lucia Melloni. The three of us work together in order to coordinate this thing, and hopefully, within two to three years, we will have answers, and I could tell you at least which parts of the theories were challenged. Or if things go extremely well which theory was challenged substantially, which is what we are hoping to achieve here.

One of the most amazing things about this project is that besides the ability or the opportunity to arbitrate between different theories and find exciting findings about the brain and consciousness, it also promotes open science and team science. I get to work with the most amazing group of postdocs and PIs, principal investigators, from whom I, as a relatively young researcher, learn all the time. We get to work with philosophers. We have the proponents of the theories and, of course, the two people with which I work regularly, Michael Pitts and Lucia Melloni, together we are trying to lead this effort. I just learn all the time from all these people. Aside from the scientific aspect of it, there is also a sociological aspect that I hope would also inspire others to work together in order to solve big questions. You can solve many interesting questions in your own lab, but sometimes there are questions that are so big that require big science and team science. I'm just happy to have the opportunity to be a part of such a project.

I: That is fantastic. That is wonderful that there is an entire system now to do the tests of these two leading theories in consciousness. The discussion about the theories actually made me consider a couple of interesting questions. One is especially related to the sensory experience and if there is a higher order. Then what about animals in sense, are animals conscious? We know that animals are self-aware, but how does that play into human consciousness?

R: That's an excellent question. I should say, by the way, that there is another project within the same initiative of TWCF that tests these two theories in animal models. There are also other projects of adversarial collaboration on other theories. One more thing about that project, it really is probably the most complicated thing I have ever done. The reason I am very happy about it is because I think that it's a new way to do science in a sense. In physics, they have been doing such adversarial collaborations. There have been some cases in cognitive neuroscience, but this is probably one of the largest scale ones in our field. I am learning so much both from the theory proponents and from the amazing postdocs that we have there and

from my collaborators, Michael Pitts and Lucia Melloni. I am a relatively young researcher. It's just I'm amazed by the wisdom of others and the enthusiasm about trying to solve the problem of consciousness. It was just important for me to also say that.

Going back to your question about animal consciousness, this is, of course, one of the greatest challenges. I spoke before about ethical considerations. Let's say that we were able to arbitrate between the theories. One crucial implication for that, once the field kind of converges on one theory of consciousness, that theory should also provide us with tools to determine which organisms are conscious and which are not. That relates to animals. That relates to AI, which we might discuss later on. It becomes more and more pressing. That relates to foetus, you know, feti, which, of course, is also highly important ethically. The importance of this question goes well beyond our intellectual curiosity, which is huge. Also ethically, it's very important. Here it's not as if I can draw a line and say from this threshold onwards these animals have consciousness and these animals do not. Bees are a good model to study consciousness and ants are a good model to study consciousness. Some say no. You have to have at least some kind of structurally developed cortex in order to have that. We don't have that criterion yet.

For some of the theories they have estimations. Integrated information theory claims that they have this kind of consciousness meter where they use a method called zip-and-zap. They give a TMS pulse. TMS is like a magnetic coil that allows you to give an electromagnetic pulse to a specific area and either activate it or inhibit it temporarily. Then you can see how that signal spreads throughout the brain. And using that method they were able to, for example, differentiate between a clinical state of consciousness. Similarly for GNW, there are experiments that try to diagnose patients based on their responsiveness to a series of stimuli and their neural activity that is evoked by it. We have some tests of consciousness. I would say that they are still under work, so we don't have something that I can, okay, give me that robot and I will tell you it is conscious or it is not conscious or that animal. If you ask me personally, given the theories I think that most of them would predict that most of the mammals should have consciousness. Again, this is something that is still undecided, I would say. There has been a manifest of consciousness researchers. I think it was the early 2000s where they explicitly stated that animals have consciousness and so on, but it's still under work, I would say.

I: Thank you for that. To move immediately on to artificial intelligence, instead of discussing if artificial intelligence is conscious, if we work together and maybe in the next few years have a theory of consciousness that is quite solid (even if it's not complete), do you think there is scope to recreate it artificially?

R: The million dollar question, right?

I: Yes.

R: I would say the following. I don't want to be a prophet and I don't want to speak to about things that I'm not 100% sure of, although who is 100% sure of anything these days. There are some things we can be certain of. There has been an interesting paper in science a few years ago written by Stanislas Dehaene and Sid Kouider and Hakwan Lau where they define a set of functions. That if an AI would have those functions they would consider it conscious. These functions relate to being able to monitor oneself to globally broadcast information and

integrate information. It's, of course, well-rooted within the theories from which they come. They say if there will be a system that is able to do all of that, then it will be conscious. Of course, they are not talking really about the qualitative aspect that we are interested in, like will it feel the redness of red. I don't know. In humans these functions may go with that experience. Does it necessarily have to be the case also for an AI? We don't know. If you ask the integrated information theory, it will tell you in order for that AI to be conscious, it has to be able to physically influence itself. A simulated mind cannot be conscious because it does not physically change itself. It cannot have an impact on itself from the intrinsic point of view. The theories are also kind of divided on that portion of an AI, but you also rightfully pointed out the problem of knowing.

You don't have to be a scientist. It's enough to watch *Westworld* in order to under the complexity here. We can imagine a robot or a zombie. If you are a philosopher there are famous zombie experiments thinking of people that look just like us and behave just like us but have no conscious experience. How would we ever be able to tell these two people apart, two types of people? It's a huge question. I don't have a good answer. I can only tell you that the search is fascinating. People have been trying to experiment with it and have been trying to think about it. As I said before, I think one of the reasons projects that try to arbitrate within theories are so important is because they might also get us closer to developing clear criteria about what is and what is not conscious.

I: Thank you very much. I know you just said you don't have a good answer, but I think that was a pretty good answer.

R: Thank you very much.

I: I think we're moving to the last question which is a little bit more philosophical. A friend who's a neuroscientist, Steven Di Costa, he, obviously, studied the brain, and he always described that it's a little bit like a visual illusion, that you are using your consciousness to understand consciousness. In so many visual tricks or illusions, you know what the trick is. I don't know, maybe there are lines that appear to be longer because of the shape of arrows. You know that is wrong, but you cannot convince yourself that it is not wrong. My question is, how does this understanding of consciousness help us or help humanity in general? Can it bring us more control over ourselves or it's just absolutely that we will be understanding ourselves without being able to change ourselves?

R: If I understand you correctly, you're not pushing the conversation towards the question of free will, right?

I: It can be free will. It could be just the way if by studying consciousness we can be more in control of ourselves. If you want to discuss free will, I'm happy with free will.

R: It's a small question, right, free will. Let's leave it to the last minute of our conversation. I think it's related. Our ability to control ourselves or to influence ourselves seems, at least, to some to presuppose the existence of free will in a way. I think you can think about it in several ways. It is true that there are visual illusions that we cannot break, but we actually have an experiment in the lab that is now ongoing and we're hoping to submit that also very soon - I'm telling you all

my secrets here - where we show that we can semantically prime people to either see more or less of a visual illusion. Some aspects of visual illusions are amenable in a way or prone to being influenced, but you're very correct that visual illusions are often given as an example that no matter how much we are aware of the illusion and how much we know about it, we still see the way it looks because our brain is structured in that way and our processing mechanisms are prone to processing information in a specific manner. I still think that our understanding of the problem enriches our ability to cope with it in a way. I'm not sure it's a problem. My fascination with the question of consciousness is not because I want to change my consciousness. I'm pretty happy with our conscious experience. I think that human beings have been able to do amazing stuff with the way they experience the world. We've also done some pretty bad stuff, but we can focus on the good side.

For me, personally, the main motivation is just the puzzle. To understand how come that neural activity, which is completely physical, how does it give rise to a quantitative experience that is uniquely mine that has a subjective character to it? It's so distinct and so clear. How can that be? How can it be like the taste of chocolate? It's a miracle. It's a wonder. I really want to understand that wonder. My motivation is not to control my demeanor or whatever better, but I do think that there is also an interesting question about the role consciousness might play in our decision-making. That is where I think free will gets into the picture. I think it's very interesting to ask how much of decisions are made consciously or deliberately and how much of them are unconsciously motivated. Try to tease those apart and ask what is the role of consciousness in these types of decisions. I think here knowledge can actually help you.

For example, when I watch TV with my kids now and I see some blunt example of a product placement where the main character is doing something with some brand new computer, I warn my kids. I tell them now they are trying to play a trick on your consciousness. They are trying to persuade you. Here I think that becoming more aware of the unconscious influences on our decisions is actually beneficial and might help us break the illusion. I belong to those that think that our decisions are not free in the sense that we can do otherwise. I'm happy about it in the sense that I wouldn't want to be a creature that is completely random, that acts in a different way when the same decision is present to him or her over and over again. I actually want to make the same decision if I'm presented with the same alternatives because I think that is what makes me consistent and that is what makes me Liad. I don't want to have some random flip coin to determine if I'm going to study this or that or maybe that person or another. Even deciding which chocolate to eat, which - as you already understood - is a major decision for me. I want to be driven by my values and my beliefs and my knowledge rather than by some enigmatic freewheel that is completely random.

In a way, I think we can be happy about our lack of freedom and understand that this is actually a new way of freedom. Acting based on your own motivation, not always we can do that. Sometimes we act and feel and ask why did I do that. But this is already a completely philosophical discussion, and there are many philosophers who would think that what I'm saying here is probably nonsense. But some philosophers I know would agree with me because I read their books, and I'm basically quoting them to you now.

I: That was absolutely great. The entire discussion has been absolutely fantastic. Please let me thank you, Professor Mudrik, for this fantastic journey into consciousness, what we know and what we don't know and what we might soon find out. Thank you.

R: Thank you so very much. It's been a pleasure.

[END OF TRANSCRIPT]