Emotions and Consciousness In Robots

Transcript of the Human-Robot Interaction Podcast Episode 21. Published on 1 July 2022 by Christoph Bartneck at:

https://www.human-robot-interaction.org/2022/07/01/emotions-and-consciousness-in-robots/

[00:00:00] **Christoph:** In the last HRI podcast episode we talked about the role of emotions for humans and robots. The interview with Douglas Campbell truly inspired me. And there was not enough time in the last episode to go into the full depth of our conversation. So here it is. Our full discussion about emotions and consciousness in robots.

[00:00:26] This is the human robot interaction podcast. I am your host, Christoph Bartneck

[00:00:46] **Douglas:** Okay. Yeah. I'm a philosopher. I'm interested in all kinds of things. My areas of research are really quite sprawling. I'm very interested in artificial intelligence. That was what really got me into philosophy in the first place. But since then, I've also got really interested in metaphysics. What the world is ultimately made often. I'm interested in altruism so why does it that altruism arrive altruism evolved and human beings.

[00:01:08] **Christoph:** Are there any fundamental differences between human emotions and robots that are able to perceive, process and express emotions?

[00:01:17] **Douglas:** I think the answer that question will be different depending on the philosophy you ask. But I think my answer will be a fairly common answer. I've subscribed to a certain theory in philosophy called functionalism. The idea there is that it's what it is about the human brain which gives rise to consciousness and thoughts and feelings and emotions is the information processing that's been performed in the brain by, the neurons or perhaps also in virtue of hormones and other things that are floating around inside the blood. So the idea is that all of that information processing has to be right in order for a thing to have emotions and feelings and experiences and consciousness. And so my suspicion very strong suspicion is it, at present any robot we have that appears to be emoting, and that appears to be able to register emotions and people in respond appropriately isn't doing that. Using the same information processing that a human being is using. They're probably using much less sophisticated information processing than we're using. And as a result, I think it's very unlikely that they have any of that going on. They're basically zombies. They may look like they have emotions and they may genuinely be able to detect emotions in people, but they themselves don't have any emotions.

[00:02:39] They couldn't sympathize with us or empathize with us at all because they couldn't feel those emotions themselves simply because they're not processing information in the right way. But that's not to say he did it in the future is as technology becomes even more sophisticated. And as we learn more and more about the brain and how to copy that simulate to the brain within , a machine. That's not to say that there couldn't one day be a robot that has emotions just like us.

[00:03:10] Christoph: But what makes feeling emotions so special?

[00:03:17] **Douglas:** I think that it's absolutely pivotal to our existence that we have at least consciousness, right? So that we have an awareness of the world. If we don't have consciousness, then it would be as if we didn't exist. From the point of view of a third party, they could see us moving around and doing things, but we simply wouldn't have a point of view at all. We wouldn't be there. We wouldn't be seeing anything or experiencing anything.

[00:03:43] Consciousnesses is crucial. So if a robot doesn't have consciousness, then there's nothing there inside it. There's no one inside looking out at the world. So that's consciousness in one aspect of human consciousness is emotion. Now how crucial is it for a conscious being to have emotion? I don't think it's completely crucial.

[00:04:08] So you can imagine this Dr. Spock, of of Star Trek, this alien who doesn't have emotion, but who's presumably conscious and fully aware of the world and able to think intelligently. I think it's perfectly possible for there to be a robot like that, something that is conscious, but that doesn't have emotions.

[00:04:28] So to that extent, I don't think that emotions are as important as consciousness, right? You must have consciousness in order for there to be a world that, you're in order for you to be seeing reality at all, but it's possible to have consciousness without emotions.

[00:04:42] Christoph: One of my other guests brought up the distinction between having emotions and feeling emotions. And she mentioned that robots can have emotions, but they don't feel them. So she can create a computer program. That can sort of the computer or the robot can sense the emotions, process it according to certain algorithms or some sort of software and express emotions. And that is what she defined as having emotions, but she distinguished that from feeling emotions and she argued that a robot cannot feel emotions. Would you think this is a good distinction?

[00:05:23] **Douglas:** It's it's certainly a perfectly valid way of using words. I think. Instead distinguish between simulating emotions and both feeling and having emotions. I would, that's how I prefer to use words, but it doesn't matter. We can use words her way if we like.

[00:05:38] So there's an absolute distinction between acting as if you feel emotions. And feeling emotions. So I can easily, at the moment we can already build it robots that have emotions in the sense that they can make, they act as if they feel them, but they don't feel them.

[00:05:55] But if she's saying that there will never be a robot that actually genuinely feels emotions, then I'm inclined to disagree with her on that. So I think we know one day we easily could have such a robot.

[00:06:07] **Christoph:** So you mentioned the right way of processing emotions or the information around it, let's say. So is your argument that it's just a question of complexity and understanding exactly how humans work. And as soon as we know how the humans work, we can model that in software. And if it is then very similar to what humans do, that will be the right way. And would that mean, then it would feel emotions.

[00:06:37] **Douglas:** Humans have a very particular evolutionary design. We've been shaped by the environment that we evolved in. So we have a whole suite of different emotions that are very human. We feel disgust. We feel shame. We feel pity. It's easy to imagine other perhaps aliens who are just as intelligent as us, but who perhaps aren't social animals like us and therefore let some of those emotions that don't perhaps have any feeling of emotion of guilt.

[00:07:06] So I think that if. A good enough job. It doesn't have to be perfect, but a good enough job of duplicating the information processing that occurs in the human brain. Then you would thereby create a robot that has our emotions. And so you would have found a right way to do it. But it's not the right way because of course there's any number of other intelligent creatures out there on the universe who would feel different emotions from us and their different way of processing information is also another right way to process information, to have emotions does it makes sense?

[00:07:40] **Christoph:** One of the kind of little interesting sentences that has been uttered about artificial intelligence is that an airplane doesn't need to flap its wings to fly. So therefore, with emotions, as long as a robot can produce output that is plausible. Would that not already be sufficient? Independently of how it is processing the information as long as it is plausible. Would that not be good enough?

[00:08:17] **Douglas:** So there was once a really popular theory in psychology and in philosophy called behaviorism, which was basically, that was precisely that, the idea was that all you need to do in order to produce consciousness or feelings or thoughts, experiences, emotions is act in the right way. As long as the behavior is correct, then the objecting question, whether it be a human or an animal or a robot has all of the consciousness that goes along with it. Nowadays that's roundly repudiated by almost everybody.

[00:08:49] There are a number of reasons for that. The one noise thought experiment as up super stoic, right? Who's a super stoic it's someone who is very, very stoical. They're actually very sensitive to pain. When they're injured and when their body is injured they feel incredible pain and they're very scared of it.

[00:09:07] They have all of these emotions of fear and terror and afterwards they're in shock. But because they're so stoical, they are able to act just as if they didn't care at all. It seems that such a person as possible, right?

[00:09:24] So the information processing that's going on inside their brain as the information processing that causes all of these emotions of terror and, pain and fear and so forth. But the behavior is just there to have a person who's, doesn't really feel pain at all.

[00:09:39] If you think that such a person is conceivable, when I do then that completely undermines the idea that it's only the behavior that man. And why should just the behavior matter? There are all of these different ways that the brain could process information that would produce exactly the same behavior.

[00:09:55] Why should that mode of information processing be completely irrelevant? Why should it be only the behavior that matters and not the information processing that's going on? It just seems like a Superficially plausible idea. But once you think about it, it's just that doesn't make any sense.

[00:10:11] **Christoph:** Hiroshi Ishiguro took the example of animals and his argument was that an animal, maybe a simple animal might not have all the. Complexities of cognition and intelligence as humans have, but they have emotions, maybe simpler emotions, and it is still an emotion and a robot could be just like a simple animal, but it would still have emotions. Would that make sense?

[00:10:40] **Douglas:** As I said before, in my view, it's not quite right to speak of processing information that the right way is if there's only one. Rather, there will be all of these different ways of processing information, obviously infinity of different ways of processing information.

[00:10:57] Of those, some will give rise to consciousness along, putting perhaps emotions in some cases and a whole lot won't. And among those that do, there'll be some that give rise to human, like emotions and others that give to, weird kind of alien emotions and others that perhaps give rise to, the emotion that your dog has when it's waiting for for dinner and you go to the fridge door and open it and then close it and don't bring out the food. I tend to believe that dogs are capable of feeling emotions. I think that they also are performing computing in the right way in order to have emotions. So the question really is whether or not the robots we've built now, whether their way of processing information puts them in that seat of different seats of ways of processing information, which give rise to emotions.

[00:11:48] And I'm just dubious. I think that it depends very much on the details. It depends entirely on the details. I could be persuaded by a computer programer, but they know, that look what we're doing is amazingly sophisticated and, look at that the way we're doing this special trickier and plausibly, this is exactly the thing that leads to genuine emotions. I could be persuaded. It's just that at the moment I haven't been. It seems to me that most of what's going on at the moment is pretty simple.

[00:12:13] **Christoph:** But are we not also confined by our inability to access the feelings and emotions in others? So I do a bit of crayfish hunting here in New Zealand. It's one of those things you do as a Kimi and we catch crayfish and we always have this lengthy debates about what is the most humane way of killing them. And the problem of course is that we don't know crayfish. If you throw them into boiling water, they make no sound. They don't have any expression that we can decipher as to it's feeling discomfort. It is just completely stoic.

[00:12:51] **Douglas:** Yep. Yep.

[00:12:54] **Christoph:** So we cannot know how it feels. So we cannot really ever know what is the right way then, because we have no way of knowing.

[00:13:03] **Douglas:** So the that's the famous problem of other minds. Every one of us knows that at least one, the conscious mind exists our own, but we can't directly detect the consciousness of anybody else. Even you for all, and, or it can be really sure of, you might be, a zombie you might not think that the simulation or possibly. This is the problem for the mines and it's an immensely important old problem and philosophy in it's what creates the so-called hard problem of consciousness.

[00:13:30] How do you try to understand consciousness when the only example that you can ever actually really detect is yours. And that means that no two scientists can see the same instance of consciousness, We all just see their own private one. It can't be studied objectively. No two people can see the same example.

[00:13:50] Consciousness is for that reason incredibly interesting and obviously a real phenomenon. They be difficult to study it's for that reason, it's hugely fascinating to philosophers. I don't have any I wish I could give you the solution right here now, but I can't. We need to distinguish between the epistemic difficulty of telling whether say a crayfish has emotions and the ontological question as to whether it does or doesn't. Even though I can't be sure that you've got consciousness, I can't actually directly see it. It's really plausible that you do. And it's really plausible with other human beings do. And it's really plausible the dogs too, as to how far it extends that's the problem of coming up with a decent theory of consciousness. So what a theory of consciousness needs to do is it needs to, if it's a functional series, it needs to precisely demarcate the seat of

things that are processing information in a way that gives rise to consciousness and tell us what kind of consciousness they're giving rise to from all of the other things that aren't.

[00:14:58] And at the moment we don't have such a theory. And so we don't have a way of drawing that line. So for that reason, everything we say about robots is a bit speculative, right? My guess based on what I know about robots at the moment is that at the moment they're not crossing the barrier.

[00:15:14] But from the fact that we don't know exactly where that barrier is at the moment, that's the epistemic problem. It doesn't follow that. There isn't a such a barrier. No there surely is. And so there really is some effect of the matter is to whether or not the crayfish is feeling emotions or not. It's just that, it's hard to know what that effect of the matter is.

[00:15:31] Christoph: I could build a very simple machine that has, let's say only one emotion pain. And if I press a button, it feels pain. And if I don't press the button, it feels happy and I could claim this is maybe like a very simple animal, maybe even less than a worm. I could argue that, look, this has emotions on the level of a worm and therefore it is processing the information similar and therefore it doesn't simulate emotions, but it literally has emotions.

[00:16:04] **Douglas:** And in presumably the way you would actually wire that up is as you would make it so that pressing of the button why is it painful? Because it's aversive. Whenever you press that button, what the machine does is it tries to work out what it's done recently, that might've caused you to press the button and then it won't do that again. Because otherwise all you're doing is you're just saying it's a pain button, right?

[00:16:25] Oh, it's a pain button. Oh, look, I press the right cause pain. What makes it a pain button? It needs to be something that makes it a pain button. I think the thing that plausibly really would turn it into a pain button is the pressing of it being aversive for the machine. And then your claim then would be that to process information in a way that's sufficient to give rise to the feeling of pain or perhaps the emotional pain. Just as having such a ability to detect and respond appropriately to aversive signals. So that's a theory of consciousness and it could be true, but probably isn't.

[00:16:59] **Christoph:** But again, if we go down the evolutionary ladder of animals to very simple animals, they don't have consciousness probably. And they are just pain, reacting machine. So on that level then would it not be the same?

[00:17:17] **Douglas:** Does the fly feel pain? Does mosquito?. Or are they just little robots? That's part of the whole great question of what the true theory of consciousness is. We don't have the correct answer. Here's one possibility, right? Even an atom is conscious, right? So that's the pan psychic possibility. No matter how your process information, you're in this consciousness. And then from there, you get to more and more restrictive theories, which draw more and more. And a stronger and stronger kind of limits on the kind of information processing that's required until you get to a really restrictive theory, which says, look, you need to be processing information pretty much exactly like a human being in order to have any consciousness, or emotions at all.

[00:17:57] Not even dogs count as having emotions. For what it's worth, which might not be very much, my money is in the middle somewhere. I'm guessing that dogs have consciousness, crayfish don't flies don't, worms definitely don't. The cutoff point is somewhere in the middle there. That's just my suspicion. I'm not sure.

[00:18:17] **Christoph:** So just to clarify, you mentioned before that you think it is possible to have consciousness without emotions and it is also possible to have emotions without consciousness.

[00:18:29] **Douglas:** No, I wouldn't say that. No. Consciousness is a little bit of a tricky word. There might be different things one means by consciousness. There's something rather strange about the idea of something that's having an unconscious emotion, right? It's like the whole what's distinctive about emotions is the way they are experienced, the way they feel for the person who's having them. What makes anger, anger? It's the kind of the way it feels to you and what makes grief grief is the way it feels to you. And if you're not conscious, Nothing feels like anything to you, right? There's no you.

[00:19:05] **Christoph**: So is it conceivable to have emotions without consciousness and your argument is that no, it doesn't really make any sense because you cannot really reason argue or even think about anything around you or your own sensation. And therefore it would just be a pain machine.

[00:19:25] **Douglas:** It would be it would be a nothingness, right? It would be pain that was experienced by nobody. What is that? That's not pain at all. Pain has to be experienced by somebody. What I'm saying is there needs to be someone who experiences it. I'm not saying that person, that thing, that experiences the pain has to then be capable of remembering their pain and conceptualizing it as pain or having detailed thoughts about it. It could just be a very simple consciousness of the pain, but it needs to be some consciousness, at least of the pain.

[00:20:00] **Christoph:** But now we're having a really difficult situation because well, this podcast was just supposed to be about emotions and emotions alone. I wasn't planning to touch on consciousness or strong AI or weak AI, but now your argument is that you cannot disentangle them.

[00:20:18] **Douglas:** You can't really disentangle them. The question as to whether or not robots can have emotions the answer can only be yes. If they can be conscious, at least on my view. It's very strange. As you said before, we can distinguish perhaps between feeling emotions and then having them, and if we are only concerned with whether they have them then yeah, of course robot is going to have emotions cause they can act as if they feel emotions. I have no problem with that claim, but if we are concerned with is whether they can feel emotions then to my mind, that's takes us instantly into consciousness territory.

[00:20:52] **Christoph:** So it becomes really this distinction between having emotions. In terms of simulation and feeling emotion seems to be at the heart of the problem here. Do I understand you correctly that your argument is that unless you can feel the emotions, you will not achieve human-like consciousness?

[00:21:18] **Douglas:** There are perhaps some people who are very emotionally stunted. Perhaps sociopaths or psychopaths, they're just kinda missing a whole raft of emotions. That the rest of us have. I'm not the expert on this, but I imagine if they're, if they can be a social person, psychopaths who are missing those emotions, then there can be other people who are missing a whole lot more emotions, and yet they can be conscious, right. They're intelligent people, but they have a very limited emotional life, perhaps almost no emotional life. No, I wouldn't claim, that you couldn't have a human mind if you didn't have emotions, because I think there are some people who don't. They really don't ever have any emotions at all, but certainly it's even if it doesn't actually happen, it's conceivable.

[00:22:04] Christoph: There have been some examples in the past of people who had certain types of brain injury that resulted in them having I would say compromise the ability to process emotions or having emotions or even feeling emotions. And they had all sorts of different trickle on effects about the ability to operate in the world. And that kind of highlighted the importance of emotions for humans to function at all. So rationality itself is largely insufficient for us to function because we're just not good enough at it.

[00:22:36] I agree with that. It'd be strange if we had all these emotions and they weren't useful for anything. So surely they are. Dr. Spock is limited and in many ways, because of his lack of emotions.

[00:22:47] So this is a good issue you raise here. Mr. Spock and the depiction of science fiction. And I'm sorry, I have to be a star Trek fan. So I have to be a little bit picky here. The Vulcans in star Trek, they have emotions. Actually quite intense emotions, but they are stoic. So they control them. So yes, he has this appearance and ability to act rationally but it still has the ability to sense it

[00:23:10] **Douglas:** This so unfortunate for philosophy. It was such a that's really good. That's really good to have. Dr. Spock as someone to talk about, so I'm not allowed to anymore. I'll take that on the chin.

[00:23:20] Christoph: But putting it aside, the question that comes back in a lot of science fiction is that whenever you have a robot like Mr. Data from also from star Trek, who does not have emotion. The writer, put them in a situation or develop his character, that his goal in life is to become more human like. And that means he wants to acquire the ability to have emotion, to feel emotions. And that's not limited to Mr. Data. A lot of characters in science fiction are written that way that emotions is the one thing that machines don't have. So therefore, when they become intelligent, their goal must be to gain this ability to be more like us. Other than vanity that we might have that we want to be liked. We want to be imitated. We want to be the superior being on the planet. Other than that, would there be any real reasons for a Mr. Data to even want to have emotions?

[00:24:26] **Douglas:** I imagined that the reason why it's such a common plot point in science fiction is because it's such an obvious plot point when you're writing a science fiction series, you need to write about something. And so it's a, that's an easy trope to explore. And it leads nicely into all kinds of dramatic situations. As to whether or not a robot would actually want to experience human emotions. So emotions are what define the peak of our lives and the absolute troughs in our lives. So when you have a wonderful, wonderful experience, that's emotions, there's happiness and joy and all these other emotions.

[00:25:10] And when you're just in the pit of despair and grief those are more emotions. To have a life without any emotions would be a very bland thing. So I can imagine that robots might having appreciated that want the good emotions, but not so much of the bad. There are emotions are very mixed bag and some of them are the most special things about our existence and some of them are the most awful things, about our existence.

[00:25:40] Christoph: Fiction responds to, of course the development in computer science and previously it was thought that if a computer can beat the world chess master, then this computer would be intelligent. Then that happened. And then we moved our goalposts to Go and then that happened and then we move it further to jeopardy and it happened. So writers of science fiction have the problem that they can no longer argue about. Look, there's a machine. It can calculate it can think. Everybody assumes that this is already happening and it is normal for it to be.

[00:26:20] And the argument is that this one thing that machines can not have and will never have is emotions. And that is what makes humans special. And therefore we can clap ourselves on the shoulder as being a superior being in terms of we have it, you don't. But is that justified? Is that really something we can be proud of?

[00:26:45] **Douglas:** There were two questions there. So when is it justified? And I would say no. It's a matter of processing information in the right way. And sooner or later we'll crack that problem just as we've corrected the others. And then we will have a machine that has emotions.

[00:26:57] So that's the first thing. The second thing is, should we. Clap ourselves on the back for that? To the extent that having beautiful emotions is that, that what makes being conscious a valuable thing. Of course, yeah, we should clip ourselves on the back for being able to have that. So it's a wonderful thing if we didn't have that life would be worthless. That's the reason why, if I was a machine that was intelligent and, but they didn't have emotions, but that learned that having emotions was this bliss, having certain emotions was this really, really positive, nice thing. Then I would probably want to have them.

[00:27:36] Christoph: Now we have got a whole religion is based on those, which is Buddhism. And the idea there is life is suffering and their goal is to free yourself from all of that, to be in this Zen style of a state where you don't feel anything you can be in the now. And that is the most desirable state to be in, which is kind of like being a robot.

[00:28:03] Douglas: I think I would prefer to be myself in a state of blissful joy rather than of nothingness,

[00:28:12] **Christoph:** Which brings me to this one question I ask all my interviewees in this series. Would you be happy to be able to turn off your emotions sometimes?

[00:28:21] **Douglas:** Emotions are very mixed bag. Often they are very helpful and are appropriate. And if you didn't have them, you wouldn't function very well. There are times when getting a bit angry is really the correct an inappropriate response when it gets you the results that are required. But there are other times when it's really not. A common example from my life is when I when there's a cupboard door at it just below head hight and I stand up sharply and to wick my heat on the corner of the cupboard door and it really hurts and I cannot help but feel angry.

[00:28:54] I think it's an evolutionary reason for that. And then the past, back in the Pleistocene, when we were evolving and when out of the blue you suddenly got wacked on the head really hard. And then probably someone had just ambushed you and was about to murder you. And the appropriate response in that scenario in the Pleistocene was just crazily angry and and then you, if you did that you might get away. But it's not appropriate in the modern day. It's not appropriate to get really angry at your cupboard door when you're the idiot who you hit on it. Yeah, I think there are many occasions like that when I would gladly turn off my emotions, because they're just not appropriate. They're not helpful. Another example is, grief, horrible grief that a parent has, when their child dies or, the grief of a, a lover who's been dumped. They're just horrible emotions and it would be wonderful to have a little switch behind your ear. You could just switch the emotion off and give yourself a break for a few hours. I would gladly do away with some of my emotions on occasions.

[00:29:52] **Christoph:** In summary, your argument is that you can have consciousness without emotions, but you can't have emotions without consciousness.

[00:30:02] **Douglas:** Yes, although I would put it that you can't feel emotions without consciousness because earlier you distinguish the having from the feeling.

[00:30:08] **Christoph:** So that's the second conclusion is there's a difference between having or simulating emotions versus experiencing them. I cannot describe it any other way as the experience of the feeling. I don't know what else I can employ to communicate this.

[00:30:26] **Douglas:** It's one of those things you can't describe to anybody who hasn't had it. It's like trying to describe the color red to someone who's never seen color. You can't. There are no words to describe feeling in consciousness except those three words, which wouldn't make sense to somebody who hadn't actually felt it.

[00:30:45] **Christoph:** So this may be our problem here of creating robots that feel emotions is that we cannot write software to do it because we cannot explicitly define or write things that describe it appropriately. And therefore we cannot teach them or make them experience it because we just lack the absolute ability to describe it.

[00:31:14] **Douglas:** That would be unfortunate if it was true. I hope it isn't true. And I don't think it's true. I think that what will happen is we keep on rapidly learning about the brain as we are, and in discovering the way that it processes information as we're going to learn that there are certain nifty really nifty tricks the brain is doing when it processes information. Which are responsible for us being able to do many of the things that even now AI at the moment, can't. Things like really creatively solving problems and having kind of serial, conscious thoughts, one after another, where we can apply logic and make inferences in a way that's also partly shaped by kind of learning patterns in the world.

[00:31:59] Once we make those discoveries and port those those learnings across into Silicon and build a machine that processes information in the same way we're suddenly gonna have on our hands a true Al. My guess is that when we do that, we're also going to see in that information processing that algorithm that we've discovered something really bizarre and exciting then, and someone's going to say, that's, what's responsible for emotions that's there. And that's going to be a really plausible theory but based on the role that it plays in helping the machines, that have this perform well.

[00:32:43] That will then help us draw the line and say, okay, oh look, human beings have so good circuitry. They have emotions or dogs, have that, circuitry, they have emotions. Oh, crayfish don't. They don't have emotions. So I think that's a way that science could overcome this epistemic barrier.

[00:33:00] Christoph: So could you consider then robotics as a tool to test how well we understood ourselves?

[00:33:11] **Douglas:** Absolutely. If you can't build it, you don't understand it. When we can build a thing that acts like us and that emotes like us and it processes that information like us. Then you understood it, right? If you think you've understood how the brain processes information and you transpose it over into Silicon and you build this machine and it doesn't act like us you know, you've made a mistake somewhere.

[00:33:34] Christoph: If our brain would be simple enough for us to understand we would be too stupid.

[00:33:41] **Douglas:** Yeah. That's a thought. A guy called Colin McGinn came up with this idea. The idea is that it's perfectly possible to understand what consciousness is. It's just that we're too stupid to do it. He came up with the term cognitively closed. There are some problems that are just too hard for us, and we know that the solutions are cognitively closed from us. He suggested that understanding consciousness as one of those. But I don't think that's plausible, among other things, you can just shave the brain. Do this tiny little bit of everything shaving of the brain, which allows you to completely map it out. It meant that all of it circuitry and and build a replica or a working model of the human brain that way. And then play with that add bits to it, perhaps toy with it in and see. So some of the modifications you make will make it dumber. Some of them will make it smarter and eventually we can build something that's really smart in that way. And then we can get it to explain it to us how consciousness works.

[00:34:38] Christoph: The last part is interesting one. It will explain itself to us. Yeah. Because again, we, since we are locked into our own mind and our own emotions, We cannot really tell if the other is working like us. So it is conceivable that we will be able to maybe create a robot that thinks and feels and can say things about it to us, but we still wouldn't be able to verify it because of this communication barrier.

[00:35:11] **Douglas:** Of course there's a sense in which thats problems completely general. We aren't to verify that, the table is in front of me because for all I know, I'm in a simulation and the table wasn't really in front of me. And consciousness is even more difficult than that. I can be pretty sure that the table was there, but I can

be somewhat less sure that your consciousness there, there's a deeper epistemic problem with consciousness than with the table, but just as with the table. In practice, we can ever reasons for being really sure that table is there.

[00:35:40] Once we've understood how the brain processes information in enough detail, we could have really, really solid grounds for believing, that a certain theory of consciousness is true. And for thinking that some things are conscious and that other things aren't I don't think that epistemic barrier is insurmountable.

[00:35:55] **Christoph:** That's an optimistic view of the world. None of us can predict the future. I don't think personally that I will live long enough to experience such a strongly AI. This will be probably something at best for our grandchildren.

[00:36:09] **Douglas:** Yeah.

[00:36:10] Christoph: Thank you so much. This has been a great interview. I learned a lot.

[00:36:13] **Douglas:** I think that was a lot of fun.