

The Sad Robot

Transcript of the Human-Robot Interaction Podcast Episode 20. Published on 1 June 2022 by Christoph Bartneck at:

<https://www.human-robot-interaction.org/2022/06/01/the-sad-robot/>

[00:00:00] **Christoph:** Emotions in robots are a frequent topic not only in science, but also in fiction. Robots, such as Mr. Data from Star Trek, try to become more human by being able to accurately perceive process and express emotions. Here's a short dialog between Mr. Data and captain Picard. Mr Data recently had an emotion chip added to his system. They are boarding a hostile spaceship.

[00:00:30] **Mr. Data:** Captain. I believe I am feeling anxiety, It is an intriguing sensation, the most distressed

[00:00:39] **Picard:** I'm sure it's a fascinating experience, but perhaps you should deactivate your emotion chip for now.

[00:00:45] **Mr. Data:** Good idea, sir. Done.

[00:00:50] **Picard:** Data, there are times that I envy you.

[00:00:53] **Christoph:** I admire Mr. Data's ability to simply turn off his emotions. This will become very useful in those sleepless nights in which my brain decides to not let me find any rest. In this episode of the human robot interaction podcast. We will discuss the role that emotions play for humans and robots.

[00:01:15] This is the human robot interaction podcast. I am your host, Christoph Bartneck.

[00:01:36] Robots are clearly able to perceive emotions of others, to some degree. It might not be perfect, but it might be at least not much worse than the accuracy of humans. We also struggle at times to decode the behavior of others. In, particularly when interacting with people from other cultures.

[00:01:57] Paul Ekman and his colleagues studied the human face already in the 1970s. And they were interested in the question if there are a set of universal, emotional facial expressions that work across all cultures. They came up with a set of basic emotions. The set is often mistaken for a set of basic emotions that humans can experience.

[00:02:24] That is not the case. It is only about facial expressions. Ekman also developed the facial action coding system. Which describe the units and their animation in a human face. We will get back to these once we talk about Androids.

Interview with Rosalind Picard

[00:02:45] **Christoph:** Could you please shortly explain what the difference between affect, emotions, mood, and feelings are?

[00:02:54] **Rosalind:** At a high level, these are all different words that describe affective phenomenon, where I use an affect with the letter a to mean phenomenon that include emotions but also other kinds of feelings.

[00:03:09] Hi, I'm Rosalind Picard. I'm a professor at the MIT media lab in Cambridge, Massachusetts, and also a co-founder chairman and chief scientist of a spin-out company called Empatica, which is Italian for empathatic.

[00:03:25] I think of affect as the larger category and emotions as more specific things like joy, anger, sadness, those are ones that have been studied a lot by emotion theorists. In side, the space of an affect, some things like the state of interest or the state of pain are not classically studied as emotion, but they are aspect of phenomenon that involve feelings. Also we think about not just the larger category of affect and the more focused, specific named emotions, but the timescale at which things happen. Emotions are usually described over a fairly short timescale. Like you chuckled and felt a moment of joy.

[00:04:14] But you can also have long-term moods and those are described. More like to think of it like weather. We might describe short-term weather, the amount of inches of rain you get, but you might also describe that the whole day was a dreary day. Even though moments of that day, the sun might have popped out. The long term description might be something like dreary or it might be something like, a spectacular, gorgeous day.

[00:04:44] And similarly mood tends to describe in an attempt to summarize the longer-term affective, a mix of things. It might be that you experienced a lot of different emotions, but overall the mood was, Hey, you're in a great mood or, oh no, you know, you're not in a good mood or you're kind of down in the dumps. That you're stressed or you're calm or that you're depressed, which can even take on a mental health meaning that may consist of more specific things.

[00:05:12] So mood, unlike emotion tends to refer to longer term descriptions that relate to feelings. And then finally the word feelings, feelings is something that really only you have access to. It relates to your internal experience. It's a combination of your thoughts, your physiology, maybe your biochemistry perhaps some parts of human experience that we don't fully understand, because we certainly do not fully understand consciousness.

[00:05:48] And usually the feelings are something that you are conscious of. Now. It doesn't mean that you have to be conscious of the feelings in order for them to have an effect on you. It's certainly possible that the little biochemical changes and physiological changes that are associated with feelings could affect you without you consciously realizing it. But generally when we talk about, what do you feel that is referring to your conscious and subjective experience.

[00:06:22] **Christoph:** And what role does affect play for humans?

[00:06:26] **Rosalind:** Well, this is an exciting area of research. When I first started working on affective computing, before I named it affective computing, I thought emotions were mainly playing a role of making us irrational. That if you were emotional, that was not a good way to be. What we've learned now is not only emotions, but also the larger affective phenomenon play really important roles, affecting our intelligence. Actually making us more intelligent, not less intelligent. Of course you can still have emotions, get totally dysregulated and out of whack. And that is still usually undesirable.

[00:07:07] But the thing we don't realize is that there's always weather. There's always an affective state. There's always some background changing affective signals and they are doing things like messaging your brain, what to pay attention to when you're in need of shifting your perception or your action.

[00:07:31] Little important decisions like whether or not you should even get out of bed in the morning are influenced significantly by your feelings. So the role affect plays is a constant role. Even when we don't notice emotions, they are playing important roles, shaping our cognition, our perception, our action, our relationships. Really just about everything.

[00:07:54] **Christoph:** And what happens to humans that are unable to experience emotions? How does their behavior change?

[00:08:01] **Rosalind:** It's interesting that some people can have emotions, but not be aware of them or think they don't even have them. I've met people like this in engineering and computer science. And in my work with people who are quite high functioning in many ways that might have another diagnosis like autism. And they'll sometimes say, Ros, I don't know what you mean by feelings. I don't think I have feelings. They still have emotion like mechanisms in many ways they can, in many cases, still experienced extreme emotions and in many cases they could still reason about motion. But there are cases where, for example, due to brain damage parts of the emotion system in the brain are not functioning in the normal way. This can impact experience of emotion as well as the roles that emotions play in shaping intelligent behavior.

[00:08:59] In these cases people can actually do things that are very irrational or not do things that they should do. Cause it's like, they're not bothered by the things that would bother a normal person. So they'll keep looping around doing the same stupid thing, not being bothered by it. Similarly we see problems, for example, when people who don't feel pain, like normal people do and they may injure themselves quite severely. Through these interesting atypical ways that some people are wired, we can learn a lot about the valuable role that emotions and feelings and many kinds of affective phenomenon play.

[00:09:36] **Christoph:** Is a human without emotions like a robot?

[00:09:39] **Rosalind:** I would not say that a human without emotions is like a robot. We're still extremely different from robots, whether or not we are showing our affective expressions or whether or not we have fully functioning emotions systems or affective systems.

[00:09:57] That said, not having fully functional emotion system does cause a lot of problems. And some of those problems are going to look like problems that robots have. When robots, for example, follow rigid rules

and can't adapt flexibly to complex unpredictable inputs, things that we think emotions help people do or have more rational behavior selection that adapts to people's emotions.

[00:10:25] There are lots of roles emotions play that when they don't play those roles, there are going to be problems. And those problems may look the same across some people and some machines. But that doesn't mean that we are like a robot.

[00:10:41] **Christoph:** Can robots read the emotions in human faces?

[00:10:45] **Rosalind:** I want to be clear that when you see a smile, you know if there's no context around that smile, that it could be just a fake smile, right? So I want to separate the facial expressions that may look like a frown or a smile or an eyebrow raise that may look like you're interested or enjoying something from the underlying emotion. What we can teach and have taught robots to do is to process the video coming in and give it a label like joy or sadness that is taught to match a label that a human would give. So if most humans would label the face as looking interested or looking bored or looking confused, then we want the robot to give it the same kind of label and robots have gotten pretty good at that.

[00:11:39] But that does not mean they're reading your underlying feelings. That said robots now can read more from human faces than most people can. It doesn't mean they're necessarily better at reading emotions cause they still aren't good at reading the context around the person. But they can do things now, like process the video of your face and see changes in the color of your skin that reflect your heart rate changes and your heart rate variability. And they can also pick up subtle movements that reflect your respiration changes. And these might indicate that you are experiencing more excitement and more stress. Even if you are keeping a neutral facial expression. So machines now can read that kind of information, even though most people cannot read that information when given only the video of your face.

[00:12:40] **Christoph:** How good are robots at extracting emotions from gestures and body language?

[00:12:46] **Rosalind:** Robots are getting pretty good at tracking where your arms and legs are. If you're crossing your arms or pointing at something or sweeping your hands in a kind of a joyful gesture. But my understanding of how well they can actually understand emotion from this is that right now, it's still very limited.

[00:13:07] In fact, I would say people are not very good at extracting emotions from just gestures and body language without also having a lot of information about the context that they can reason about. And in some cases also it's really important to have information about the culture and the personality. Some people in some situations will make a movement or a gesture that will mean something very different in one culture than in another.

[00:13:38] And while machines can be taught what the nod or the shake is likely to mean in these different cultures and situations. It's still doesn't tell you what the underlying feeling is that was likely to produce it

[00:13:52] **Christoph:** Often it is less important what is said, but how it is said. How accurate are robots in perceiving emotions from speech or text?

[00:14:02] **Rosalind:** The interpretation of text, as people know, when they've talked with chatbots is improving rapidly. Machines are starting to show much more progress in simply understanding what words were spoken. They still do not have good understanding of meaning or context. And because of that, even when they can interpret things, like that the voice sounds like it may be a little more tired or the person's sick or that the voice sounds more energetic, which may be more associated with excitement or joy.

[00:14:36] And while they can process more and more of that vocal tone, it's still, doesn't tell you for sure what the person is feeling. So all of this information, while increasingly the machine can start to label it more like a human listener might label it. We human listeners know that it's not telling you the underlying emotion that is truly giving rise to it.

[00:15:06] It is however, allowing you to, in a context, start to make more informed guesses about what that emotion might be. So when it's all put together by a person it can give you a pretty good guess and machines are getting closer to recognizing the pieces of that. But still they're far I think from understanding what we're really feeling.

[00:15:32] **Christoph:** What other methods are used to measure the emotional state humans.

[00:15:38] **Rosalind:** We in our lab have worked a lot on multimodal methods. Putting together the face, the voice, the posture, the gesture, the actions that you're choosing, the words that you're choosing, what you're

clicking on, what you're actually choosing to do versus not do. We also have created a number of wearable devices that if you choose to opt in and where these can also help capture physiological changes, that could be really important for the way that affect interacts with your health. And when we put all of that together, we're actually getting pretty good at doing things like telling if you're having a good morning or evening or a not so good mood morning or evening, or whether you're in a state that's more calm or a state that's more stressed. We're also actually able now to even use this multimodal long-term pattern data to forecast if tomorrow you're likely to be more stressed and more calm and in a better and more positive mood or in a more sad mood.

[00:16:48] **Christoph:** Detecting extreme expressions of emotions might be possible. But what about subtle expression?

[00:16:54] **Rosalind:** It's interesting because machines are actually quite good at detecting several nuanced things that people usually miss. For example, high-speed flashes of an expression on the face sometimes called micro-expressions can be picked up by a high frame rate camera better than most people can pick them up unless people are trained to recognize them.

[00:17:17] Also, there are subtle differences in things like the symmetry of the onset of a smile versus its offset. The dynamic trajectory, how slowly it grows versus how rapidly sharply it's pulled up the slower one tends to be associated with a smile grilling with real joy. The more knee-jerk type smile, maybe more of kind of like, I'm just kind of smiling, nice kind of smile.

[00:17:39] And these subtle differences are actually better picked up in many ways by machines. Again, that does not mean that they are better at detecting your subtle emotional state changes, which might be better predicted by understanding you and a situation you're in. So the very best understanding I think comes from putting all of these together from interpreting the context, reading multiple modalities of data forecasting, what one might expect you to experience given knowledge about a situation and then recognizing in real time, how you actually are behaving and expressing in that situation.

[00:18:23]

Interview with Lola Canamero

[00:18:29] **Christoph:** Once the robot is aware of the emotions of others. It also needs to have an understanding of the events and objects around it. They all might influence how the robot feels. Which brings us to the topic of how robots model and reason about their own emotions and those of others. We have the pleasure to have Lola Canamero on the show. Lola, who are you?

[00:18:58] **Lola:** Oh, who am I. Uh, that's a very philosophical question. I'm currently professor at the university of called Cergy Paris university. I do a lot of research on modeling emotion in a autonomous and social robots.

[00:19:14] **Christoph:** How do robots process emotions?

[00:19:17] **Lola:** It depends on what you mean by process. That might mean that they detect emotions in people and they try to process those signals and respond to them. That might mean that they might have something a bit similar to an emotional system, in the controller, something that is. Functionally equivalent.

[00:19:38] And it might also mean that they produce as some sort of emotionally color behavior. So what I do is mostly the second thing. Is model something of it, a can to an emotional system, like, emotions in the brain or, and in the body. But I have an approach of embodied artificial intelligence. So I am very interested in the embodiment aspects of emotion. So how they happen through the body. So I try to simulate some sort of physiology related to motivation as well that makes the robots have their own needs and their own motivation to do things autonomously and then use very low level mechanisms like simulate hormones, when we're in different emotional states.

[00:20:23] Different hormones affect differently, everything pretty much. So our attention, our memory, even the immune system, our movements, how we perceive things, our focus, my robots will pay attention differently to different things, depending on their motivational state what's happening in the environment that made them had some sort of emotional response. Like they might be trying to avoid a danger and then they will pay attention to specific stimuli rather than others and have more like faster muscular action ecetera. So that's what I would call the, embodied approach to model emotions. Other people do other things and people would model how we reason when we're in under different emotional states by thinking about how we think about things. How we should respond, how our robots should respond to people, for example, or to a situation

under different emotional states, by retrospectively how we respond ourselves and then putting rules, for example.

[00:21:24] **Christoph:** Would it be not just enough if the robot has six emotions, you just have six variables and each of them has a value. Would that not be good enough?

[00:21:36] **Lola:** It depends on what you want to do. For maybe for a game that's enough. But if you want robots to have a bit more subtle behavior, and that interact with humans who are much more sophisticated than that, and not sufficient, we are. Not very often in a clear emotional state. Most of the time we cannot say we're in this specific state, we have many more than six emotions.

[00:22:00] **Christoph:** So how many emotions, how many different emotions should a robot have?

[00:22:06] **Lola:** Haha. Well, again, so it depends on the application, but the way I think could be approached and is one of the ways I approach it is instead of having specific emotions. You can also model what is called the dimensions of emotions. So properties underlying, all emotions. There's one of them that is very widely used.

[00:22:26] Also robotics is arousal. So it's the level of tension. When we're angry. We have high arousal where our body is very tense and we're ready to act with strong intensity. Sadness is the opposite. And another dimension is called valence and that pictures, the positive or negative responses that we have. We perceive things as something being positive that we can pursue or negative.

[00:22:52] Another dimension that is also very important is how much control we have over a situation. And that is also defined different emotion. If we have very little control, we will probably avoid the situation and if we have a lot of control of the situation will do something about it. So using those three dimensions we can imagine a space and we can walk around that dimensional space. So that will be our emotional state. specific points correspond to those basic emotions that you said. But what is interesting here is that we can easily transition from one state to another one by modeling emotions like that. Then we have a huge palette of emotion that we can model in a robot.

[00:23:40] **Christoph:** So these models of emotion are essentially a form of data representation. The question that rises is do robots need to have a data representation similar to that of humans? Or could it be something different? Would that work just as well? You mentioned before hormones modeling Hormones Well, that's not part of these categories or dimensions. So how human does an emotion model need to be?

[00:24:14] **Lola:** Hormones are mechanisms underlying both dimensions and categories. So it's one of the mechanisms that our brain and our body use to produce the emotions. When you're angry, you behave in a specific way, and you have expressions that kind of signal that you are not very willing to interact and use numeric parameters in a way that have similar dynamics on the memory of the robot or the attention. So make narrow down the attention or make it focus on specific things or the memory of the robot is going to remember some events that are relevant for that sort of situation, rather than others. You can do it in many different ways, but what is important if you want humans to perceive that as an emotion, if you want to actually exploit the functions of emotion, the benefit that emotions have is to capture in your algorithms key properties and give features of emotions.

[00:25:23] So then you can take close inspiration from biology as I do, for example, or you can take a more algorithmic approach and use rules. It really depends on your background, what your robot gonna be used for? What sort of data you have.

[00:25:41] **Christoph:** What still the question arises, whether. It will ever be possible to convincingly process emotions other than through an imitation of humans. And old debate in artificial intelligence where, somebody said that an airplane doesn't have to flap its wing in order to fly.

[00:26:00] **Lola:** Yeah.

[00:26:01] **Christoph:** So the questionnaire comes back, does a robot need to imitate humans to have a chance to have a convincing emotion processing or it doesn't matter. Any kind of algorithm, any kind of data manipulation, as long as it looks convincing. It's good enough?

[00:26:19] **Lola:** You don't need to imitate human emotions. If you want the robot to have those emotions for interaction with people, you need to do something that speaks to us. Doesn't have to be exactly our or emotions, but it cannot be any algorithm. It has to be something that conveys the intention of the robot and conveys the situation in which the robot is and can be very simple signals.

[00:26:44] Actually, the way we process emotions is we use all sort of signals and some of them can be very low level, very simple, like purpose of movement. So it doesn't need to be the same kind of postures we have, but something that is going to trigger in our brain or there's something scary this robot is trying to respond to, or this robot is trying to communicate something to me.

[00:27:08] So, no, it cannot be an algorithm. Also know emotions are not unique to humans, so other animals have emotions, but emotions have many many components. And again, if we take an evolutionary perspective, you can look at emotions evolve, and they probably started with some of the most basic elements that we still have, like bodily reactions, avoiding, dangerous.

[00:27:34] Then our nervous systems complexified then emotions became more and more complex. And in us, they're very complex, they encompass pretty much all our cognitive system, our motor system, even at the way our organs functions. We're conscious. We have this kind of property and therefore we're aware of our own emotions.

[00:27:56] **Christoph:** Consciousness. There it is. I did not want to open this can of worms. But that doesn't seem to be a way around it. I will have to come back to this.

[00:28:08] So here's a really deep and philosophical question. And since you have a degree philosophy, I think you might be able to appreciate it. Do you think that robots actually have emotions?

[00:28:21] **Lola:** Haha, that's a very good question. These robots that have emotions are actually more scientific models. So they have a system that in a very limited way works in, a similar way to human or other animal emotions. So they have an emotional system. now I wouldn't say it is similar to us or they can have feelings like we do. No, they can have robot emotions, but that doesn't mean that they have our emotions.

[00:28:54] **Christoph:** But when you say they can't feel, I could easily program the robot to have a variable called pain and I can manipulate that variable and it can be in a pain state or it can be in a happy state. I can send it to heaven if I so pleasure. So how is that different from me feeling.

[00:29:15] **Lola:** We still don't understand feelings in human. A robot is a machine. That we can disconnect and run out of batteries and we charge the batteries and it starts running again. So we are very different type of creatures.

[00:29:29] **Christoph:** But people would argue that humans are also just complicated machines and they also have an off switch.

[00:29:35] **Lola:** Yeah. In a, in a way, in a way

[00:29:37] **Christoph:** The on switch is a little bit more complicated, but the off switch is pretty clear.

[00:29:41] **Lola:** You can activate the off suite of someone that would not be a good thing to do but you can turn it on again.

[00:29:47] **Christoph:** Still the question comes back to why would you argue that robots cannot feel when I can implement an information structure that can model that.

[00:29:58] **Lola:** I don't think you will be modeling the complexity of human feeling. You will be programming a variable that will behave in specific ways. And you might be trying to answer a specific question that maybe you have her from neuroscience. Or you might pose a question that might be of interest to them, but is an algorithm is a program is more, we're very different kind of entities.

[00:30:23] **Christoph:** But if it would just be a level of complexity, meaning that if I make it complicated enough, would that then qualify as feeling?

[00:30:32] **Lola:** No, it's not a matter of quantitative complexity, but it is, we still cannot explain feelings in humans. So you're guess you're asking me, can we model consciousness? We still don't know what it is. I don't know to what extent we can model. We cannot model the full complexity. We don't know how to model it.

[00:30:51] **Christoph:** Okay. Let's agree on the simplest organism that we have verified have some sort of nervous system that we have observed some form of emotional expression. I'm not sure which one that is, but let's say the simplest one. And let's say we've been able to model that in software reasonably completely. Would that then qualify as feeling?

[00:31:12] **Lola:** No, that will qualify as an emotional reaction, an emotional response, for example, the fact that it contracts his body or expands the body or something given a stimulus. We can model that. that doesn't qualify the way we define feelings. So I think what you after is that, feelings are these unique characteristic of humans. And if we can replicate it. We don't know if it is unique to us either. But that is for sure. One of the things that we still cannot model in rules because we still don't understand them. That's my point.

[00:31:47] **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.

[00:32:20] **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? 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.

[00:32:53] **Christoph:** This is Douglas Campbell from the University of Canterbury.

[00:32:58] **Douglas:** 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.

[00:33:20] **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?

[00:33:38] **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 you 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.

[00:34:18] 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.

[00:34:37] **Christoph:** If we don't really understand it we might also be in a situation where we have already modeled it successfully and we just don't get it ourselves we don't understand it.

[00:34:45] **Lola:** I think we would have realized that if we had model that because you would have seen some sort of responses that you could observe you could measure and you have observed like a qualitative change in.

[00:35:00] **Christoph:** I could, for example, have a robot that if it's happy, it drives around faster. And

[00:35:04] **Lola:** yeah, yeah. All that we can model. Yes, yes. We can model that. That, yeah, that is,

[00:35:09] **Christoph:** So does that mean that this robot can feel sadness and can feel happiness?

[00:35:13] **Lola:** It can have behavioral manifestations similar to the ones we have, biological systems have. So it is, again, this distinction that there are behavioral manifestations, there are physiological manifestations, cognitive manifestations expressive. if a nervous and has the property of having consciousness, then they appear there as well. So they encompass the entire nervous system. Now if you make a never system that is

conscious, then you'll probably be able to involve emotions there somehow. But we're not able to do that. Yet or ever. I don't know.

[00:35:52] **Christoph:** So are you saying that a prerequisite for emotions is consciousness?

[00:35:57] **Lola:** It's just because we have both, we have consciousness and we have motions. That is a feature definitely of human emotions, but that is not necessarily a feature of emotions. So you can have emotions without having consciousness. Other animals do. In different impairments you might lose some cognitive capacities, but you still have emotional responses. It's not a binary thing. To have it or not. We humans have conscious awareness of our emotion and these subjective feelings. But that is not a prerequisite to emotions.

[00:36:31] **Christoph:** Douglas had a different point of view.

[00:36:33] **Douglas:** 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.

[00:37:08] **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.

[00:37:27] **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.

[00:38:02] **Christoph:** Now back to Lola.

[00:38:04] The argument, as I understand it from you, is that look, emotions are complex. We don't fully understand it. We don't understand what feelings are in humans completely. And because we don't understand it, we couldn't tell if a computer does or does not have emotions or does have feelings does not have

[00:38:21] **Lola:** Oh no, we could,

[00:38:22] **Christoph:** I also a sense and this is just a feeling for my side. And please correct me if I'm wrong. A more principle hesitation to do this. Is there some, maybe some sort of belief in you that you would like to share with us?

[00:38:37] **Lola:** So I did not say that we could not recognize. But we cannot model it. We don't know how to model it.

[00:38:44] **Christoph:** If you would've done it. If you would say my model now is good and I'm satisfied with it, would that be enough to say they have feelings?

[00:38:54] **Lola:** I am not trying to replicate a human at all. I'm not trying to replicate human emotions. I'm actually very far at the other end, I'm trying to model very simple systems because I think we pay too much attention to this high level cognitive elements of emotions.

[00:39:10] And I'm trying to understand, how maybe the very basic elements can give rise to a lot of complexity. If we understand that maybe it will be perhaps a lot easier to develop maybe even treatments for affective disorders, or that could be based on other things rather than putting so much emphasis on cognitive therapies or feelings and maybe something more. I am trying to understand how very, very simple things can create already a lot of complexity and that is, the basis of more complex emotions. But if we start trying to understand everything from the top, all the complexity, we will not understand the contributions of the different elements.

[00:39:53] I'm really not trying to model consciousness. I'm trying to quite the opposite. I have even modeled that I only use arousal or only use valence and see how much emotional behavior, responses you can already

program your robot with just one dimension, then just another dimension. Then you combine them and you have more complexity, etc. So incrementally, but really from the very very beginning, we could say the very bottom.

[00:40:26]

Interview with Hiroshi Ishiguro

[00:40:35] **Christoph:** Robots do process emotions. And if you ask them how angry they are. So you can simply look up the value of that variable to give you the answer. But if a robot is in a certain emotional state, does this influence its actions? The most obvious action has to express its emotions. I have the pleasure to talk with Hiroshi Ishiguro about this topic. First of all, Hiroshi Ishiguro I would like to welcome you to the podcast and I was wondering if you could introduce yourself to our listeners.

[00:41:10] **Ishiguro:** My name is Hiroshi Ishiguro. I'm a roboticist. I'm developing the many kind of interactive robots. And basically my purpose is to understand what human is by creating various interactive robots.

[00:41:23] **Christoph:** How can robots express emotions?

[00:41:29] **Ishiguro:** Robot can use the multiple modalities in the facial expression, is one of them and the voice also, and gestures, right? So the robot can combine the several modalities to express the emotions.

[00:41:41] **Christoph:** And do robots have to mimic human expressions or could they also use more abstract, maybe even robot specific indicators?

[00:41:51] **Ishiguro:** That's interesting question. My approach is it's a kind of a scientific approach, by creating a very human like robot, we are trying to understand the meaning of human emotional expressions. In the daily possibility we may accept the robot as a kind of our partners and that robot is going to be more friendly for us. And then, we may expect the robot is going to have own emotional and expressions. So I think that is a possibility. It's quite, interesting to consider that the robot emotions.

[00:42:23] **Christoph:** And what is the advantage of using nonhuman like expressions?

[00:42:29] **Ishiguro:** It's better to have more informations, right? Comparing with other animals, we have a very rich emotional expressions, but I guess still, it is limited. And. The new features, we gonna use the more Technologies, for example, now I'm running the big project for the avatars. So avatar can have a more rich facial expressions.

[00:42:49] The avatar can have the more modalities comparing with the human. Then, the possibilities in a avatar is going to be more how can say the express than the humans. They communicate more deeply communicate with the people, many kind of people. So that is a possibility, I think.

[00:43:06] **Christoph:** You've built many great robots, and one of them is your Geminoid series. How good are these geminoids at expressing emotions.

[00:43:14] **Ishiguro:** Unfortunately, Geminoid is my copy and myself, and I'm not so good for expressing the rich emotions. The Geminoid facial expression is not so rich. Like me. We are are creating different types of Geminoid for the different people and they have more better facial expressions and gestures. The other Geminoid is good for expressing the emotions.

[00:43:40] **Christoph:** Do your avatar robots that you're currently building have more or better expressions?

[00:43:45] **Ishiguro:** Not yet we are improving that. That is our new challenge. If we have avatars that we can walk in real world without any constraints from the bodies. We can be free and we can walk anywhere anytimes by using avatars. To realize that kind of avatar societies, we are developing much better avatars right now,

[00:44:07] **Christoph:** And these avatars that you talk about are these robots or are these screen based characters?

[00:44:13] **Ishiguro:** Both of them. In order to develop the avatar market, the computer graphic screen based avatar is the more convenient. Robot can have a more human strong feeling of a presence. Last year I have started my own venture companies and the company's strategy is to develop the real market by using a computer graphics agent. After developing the market, we wanna use the robot avatars for the particular purpose and situations.

[00:44:44] **Christoph:** Humans use the are muscles in their face and in their body to move and express emotions. What type of actuators do robots use?

[00:44:53] **Ishiguro:** Currently we are using the pneumatic actuators for the very heuristic human, like and our Android, but this is of course, the, there is a limitations, right? So it's better to have more human like muscles, but that is a bottleneck. Another choice is direct to drive motor is quite flexible and it's good for mimicking the human muscle movement, but it's motor. And is not enough. We have to develop the more human like muscle. That is a very big challenge.

[00:45:26] **Christoph:** So these actuators are then attached to a skin. So they pull on a skin. Is that the basic mechanism?

[00:45:34] **Ishiguro:** For the facial expressions, we are using a small pneumatic actuators pulling the the skins. Of course it's better to have more human like muscles, for having the more sophisticated facial expressions.

[00:45:48] **Christoph:** How many of those actuators are in one of your robotic heads?

[00:45:51] **Ishiguro:** For the head, I think we are using around 20 probably for the bodies we are using 50 or 60 actuators.

[00:46:02] **Christoph:** These Androids or Geminoids are copies of humans. You also have robotic copy of yourself. Who is better at understanding and expressing emotions, you or your robotic double?

[00:46:18] **Ishiguro:** Probably that my robotic double is better because I'm a Japanese. I'm shy to express the rich, facial expression and. The Geminoid robot, it's a robot. And according to the computer program, the robot can express very rich facial expressions and some other emotional expressions.

[00:46:38] **Christoph:** While the Japanese culture, of course have more subtle ways of expressing emotions. Still Japanese people are very much capable of picking up on them to understand them.

[00:46:49] **Ishiguro:** Yeah, exactly. We are quite sensitive to the very small, the expressions like the eye contact or, the gazing patterns of eyes.

[00:46:57] **Christoph:** You said that you're not very good at, reading emotions because they are so subtle, but then again, you are also trained to read them and your robots probably too. Would it be more difficult to read and express emotions in Japan and then compared to Italy?

[00:47:13] **Ishiguro:** And probably it's quite difficult. Maybe we need to have a very big data set, but unfortunately, we do not have the big project for the, emotional recognitions. The companies is going to have that kind of the facial expression, the deep learning tools by gathering a lot of Japanese emotional expression, facial expression.

[00:47:34] **Christoph:** At the beginning, you mentioned that your job is to understand what it means to be human. Do you think that robots have emotions?

[00:47:44] **Ishiguro:** We don't know what emotion is. The human emotion is a mysteries. We don't have exact understanding about the human emotions. But we have many kind of knowledge about the emotions and then the Android and the robot can mimic the human emotions.

[00:48:00] Before that we need to understand exactly what the emotion is, right? We can do the both things simultaneously. We wanna develop the robots that can mimic the human, emotional expressions, but at the same times we may some the deeper knowledge about the human emotions through the develop the robot.

[00:48:21] So my answer is. Probably yes, someday. So we may install a kind of a emotion to the robots. I hope so.

[00:48:31] **Christoph:** So do you think that there's a difference between a robot having, let's say some variables that store information about its emotional state and the human feeling an emotion? Is there a conceptual difference?

[00:48:46] **Ishiguro:** I don't think so. Robot intelligence and human intelligence is going to be a very similar. Same things happen for the emotion. Human is a kind of a molecular machines. The robot is different, but still robot is also the machine. And if we develop the complicated systems, then we can install the emotions. We can install the intelligence for the of them. So my hypothesis they are not so different. I think.

[00:49:13] **Christoph:** So you mentioned that maybe the robots we have today are a little bit simple . But would it be fair to say that they have simple emotions?

[00:49:22] **Ishiguro:** The researcher for the consciousness, they believe that, the simple robot can have a simple consciousness. In a complicated human, it is going to have a more complicated consciousness and same things happens for the emotions. The animals, they have a kind of emotions, but their emotions are so simple. their, is not so complicated, like a human. So my answer is yes.

[00:49:46]

Interview with Douglas Campbell

[00:50:03] **Christoph:** Douglas. What do you think about this issue?

[00:50:07] **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.

[00:50:24] 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.

[00:51:15] 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 programmer, 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.

[00:51:40] **Christoph:** Are there any fundamental differences between human emotions and robots that are able to perceive, process and express emotions?

[00:51:49] **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.

[00:53:11] And so they, 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.

[00:53:42] **Christoph:** But what makes feeling emotions so special?

[00:53:49] **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.

[00:54:15] Consciousness 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.

[00:54:40] So you can imagine this Dr. Spock, 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.

[00:55:00] 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.

[00:55:14] **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?

[00:55:55] **Douglas:** 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.

[00:56:10] 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.

[00:56:27] 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.

[00:56:39] **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.

[00:57:09] **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.

[00:57:38] 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 make sense?

[00:58:12] **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?

[00:58:49] **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.

[00:59:21] 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, 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.

[00:59:39] 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?

[00:59:56] 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.

[01:00:11] 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.

[01:00:27] 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.

[01:00:43] **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.

[01:01:21] **Douglas:** Yep. Yep.

[01:01:24] **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.

[01:01:33] **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.

[01:02:00] 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.

[01:02:20] 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.

[01:03:28] 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.

[01:03:44] 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.

[01:04:01] **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.

[01:04:18] **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.

[01:04:53] **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?

[01:05:15] **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.

[01:06:02] **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.

[01:06:34] 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.

[01:06:45] 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

[01:07:07] **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.

[01:07:17] **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?

[01:08:23] **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.

[01:09:08] 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.

[01:09:38] **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.

[01:10:18] 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?

[01:10:39] **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.

[01:10:51] 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.

[01:11:31] **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.

[01:11:57] **Douglas:** I think I would prefer to be myself in a state of blissful joy rather than of nothingness.

[01:12:07] **Christoph:** In summary, your argument is that you can have consciousness without emotions, but you can't have emotions without consciousness.

[01:12:17] **Douglas:** Yes, although I would put it that you can't feel emotions without consciousness because earlier you distinguish the having from the feeling.

[01:12:23] **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.

[01:12:42] **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.

[01:13:00] **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.

[01:13:29] **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.

[01:14:14] 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 AI. 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.

[01:14:58] 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.

[01:15:15] **Christoph:** So could you consider then robotics as a tool to test how well we understood ourselves?

[01:15:26] **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.

[01:15:49] **Christoph:** If our brain would be simple enough for us to understand we would be too stupid.

[01:15:56] **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.

[01:16:53] **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.

[01:17:27] **Douglas:** Of course there's a sense in which that's 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.

[01:17:55] 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.

[01:18:10] **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?

[01:18:19] **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 height 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.

[01:18:51] 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.

[01:19:49] **Ishiguro:** I am good for controlling my emotions. I'm using my emotions for the positive ways. And even if I'm feeling some sadness, I think that is meaningful. I don't think I need to turn off my emotion. So emotion is always important. I think so.

[01:20:04] **Lola:** Emotions is something we don't have a lot of control on. Because they're, evolutionary, there old systems that are very pervasive. One of the things that we need to develop over time is emotion, regulation and control of over our own emotions. And sometimes yes, they can be more intense than we would like, or more intense than the situation requires. But that's that's learning. I would not like to switch them off. To control them better. And that's lifelong process of learning.

[01:20:40] **Rosalind:** I'm happy to have them and just develop skills and learning to regulate them. Not letting them run the show, exercising one of the great gifts of being human, which is that we can be aware of them. We can work with them.

[01:20:57] We can recognize when they're increasing or decreasing and changing. And we can both enjoy them and appreciate them. And sometimes regulate them, manage them, put them in their place. But in many cases, bring out and engage more in the things that give rise to positive emotions and savor those moments savor those things in life that bring joy.

[01:21:22] And especially after this time of pandemic, I think people need to do that a lot more. We need to pay attention to the little things that bring people joy. If somebody out there is not feeling that joy that can perhaps be encouraged by looking for ways to give away that joy, to do things for others, to elicit that joy. And then look for those good moments and savor them. And that can be a wonderfully positive gift.