Can Robots Learn to Talk?

Hertzberg, Lars

Published in:
WITTGENSTEINIAN (adj.)

Published: 01/01/2020

Document Version
Accepted author manuscript

Document License
Publisher rights policy

Can Robots Learn to Talk?

Lars Hertzberg

We have all seen or read about talking robots. They will utter greetings, make small talk, answer questions up to a point, and so on. The question is: how do robots acquire the capacity to talk? Can we think of robots as having *learnt* to talk, or are we to think about their coming to talk in a different way?

What I am interested in is not technology and its limits, but rather conceptual issues: what concerns me is the way reflecting on the question whether robots can learn to talk might throw light on *what is involved in learning to talk*, whether for a robot or a human being. By a robot we usually mean an entity all of whose functioning is directly or indirectly controlled by programming. (The control is indirect when part of the programme consists in instructions to receive inputs from the surroundings of the robot and apply them in its functioning – what is called "machine learning".) Beginning with Descartes, people have tended to fend off claims that robots may one day be indistinguishable from humans in their capacities by invoking the complexity of human cognitive abilities, the power of human beings to respond to novel situations in innovative ways, etc. This line of thought is often countered by the argument that there may be no limits to what robot technology may achieve, if not now, then in a distant future. I for one find the back-and-forth of this debate unsatisfactory. As I wish to argue, for a better understanding of these issues, we need to undertake a deep revision of the categories in which we tend to think about them. On the whole, I have a sense that discussions of how robots are like or unlike human beings should focus more on human beings and less on robots than they often do.[1]

I will start by discussing an experiment in which an attempt is made to model teaching a robot to speak on a child's learning to speak. The experimental design reveals the assumption that learning to speak is based on learning to use words to refer to objects, an assumption which seems to form a recurrent theme in Western thought about language, being expressed, among others, by St Augustine in his *Confessions*. In criticizing Augustine's conception, Wittgenstein introduces the so called builders' game, which illustrates a case in which speaking is limited to referring to objects. He wishes to show how impoverished such a language would be, a critique which is later deepened by Rush Rhees, in his arguing that learning to speak is radically different from learning to play a game. What is missing in playing a game, or in simple forms of word use such as verbal routines or skills – forms of activity which could be simulated on a robot – is the dimension of words becoming an

[1] For an essay pursuing similar lines of thought to those developed here, see Ondrej Beran, "An Attitude towards an Artificial Soul? Responses to the 'Nazi Chatbot'", *Philosophical Investigations* 41 (2018), pp. 42-69.

expression of the speaker, being given a place in her life. In the case of the robot, this dimension is lacking.

1. Is it all about objects?

The inspiration for this essay came from reading an article in *The Guardian* by Alex Beard, called "How babies learn – and why robots can't compete".[2] It was about Deb Roy, a robotics expert at MIT, married to a professor of speech pathology. In 2005, Roy initiated a research project called "The Human Speechome Project", which consisted in setting up a video recording system in their home, which recorded virtually every moment in their son's life from the moment he came home from the maternity ward until he was around three. The *Guardian* article was written 13 years after the start of the project.

This material, if treated with understanding, would seem to be an amazing source of insight into child development. However, as far as I have been able to make out, surprisingly little has come out of the project in that regard. One reason for the paucity of outcomes may be that Roy wanted to limit himself to results that could be given a quantitative treatment; that is, results that did not require trying to make sense of what went on in various interactive situations. One thing Roy studied was in which part of the house various words were the most frequently used. The words "bye bye", for instance, were most frequently heard in the hallway.[3]

In fact, the real motivation for the project was not the study of child development as such, but the possibility of using these observations in teaching robots to speak. The idea was to deal with a robot in the same way one dealt with the child in situations involving speech, in the hope that the robot like the child would pick up speech in this way. These hopes were frustrated however.

Roy used a robot he had constructed:

> Using voice recognition and pattern-analysing algorithms, Roy had painstakingly taught Toco to distinguish words and concepts within the maelstrom of everyday speech. Where previously computers learned language digitally, understanding words in relation to other words, Roy's breakthrough was to create a machine that understood their relationship to objects. Asked to pick out the red ball among a range of physical items, Toco could do it.

[2] Also in Alex Beard, *Natural Born Learners: Our Incredible Capacity to Learn and How We can Harness it* (London: Weidenfeld & Nicolson, 2018), pp. 42-45, 62-66.
[3] https://www.ted.com/talks/deb_roy_the_birth_of_a_word?language=en. Accessed 19 January, 2019.

Now this passage raises a host of questions, some of which I will get back to later. In any case, Roy gradually gave up on his ambition. As put by the journalist,

> Watching his son, Roy had been blown away by 'the incredible sophistication of what a language learner in the flesh actually looks like and does'. Infant humans didn't only regurgitate; they created, made new meaning, shared feelings… He had discovered that human learning was communal and interactive. For a robot, the acquisition of language was abstract and formulaic. For us, it was embodied, emotive, subjective, quivering with life.

Well, Roy must have known to start with, as we all do, that human learning is communal and interactive. His disappointment may have lain in the fact that he found the processes of learning did not open themselves to quantitative treatment. The project dissolves in a haze of adjectives: "embodied, emotive, quivering with life". Anything that cannot be measured, it appears, is shrouded in mystery.

What I find philosophically interesting is Roy's ambition and the thinking that lies behind it. The aim of Roy's project seems to have been to get robots to connect words with objects. This is what learning to speak is, he seems to have assumed, and this is what he failed to achieve by using the methods one uses with children. Roy recounts the following episode:

> The first time his son uttered something that wasn't just babble, Roy was sitting with him looking at pictures. "He said 'fah'", Roy explained, "but he was actually clearly referring to a fish on the wall that we were both looking at. The way I knew it was not just coincidence was that right after he looked at it and said it, he turned to me. And he had this kind of look, like a cartoon lightbulb going off – an 'Ah, now I get it' kind of look.

This episode is interesting for various reasons. It is revealing that Roy regards this episode in terms of *referring*. He fails to see that applying this term to the situation described is highly problematic. Referring is something we *do* in the course of communicative interaction. It is something I mean for my interlocutor to take in a certain way; something I, as speaker, take responsibility for, for instance, in the context of saying something true or false. Furthermore, we refer to something under some *aspect*: one might refer to the fish picture, say, as a fish, or as a fish picture, or as this particular object, or as a picture, or as an object of a certain colour, or as a funny-looking object, or as an object hanging on the wall, etc. But in the situation described by Roy the context for singling out some particular aspect is missing. The idea that the child might be referring to the fish picture as *this* rather than *that* can get no foothold. For it to do so, a conversational context would be required.

In short, the gesture Roy attributes to his son is one that in itself presupposes membership in a community of speakers.

It is significant that, until what Roy could imagine as a case of *referring* presented itself, he thought of what his son produced as "just babble" – not a very sophisticated category to use for someone who is interested in studying how children learn to talk. (It should be admitted that this formulation may be that of the reporter rather than Roy's own.) Roy saw the process through a filter, in which he registered only occurrences that seemed to him to confirm certain preconceived notions of what it is to become a speaker. Along similar lines, there is talk about the robot Toco as acquiring the ability to pick out words (and, yes, concepts!) in "the maelstrom of everyday speech" – words, it seems, are what gives structure to the inchoate stream of speech sounds; what needs to be discerned against a background of "semantic noise", as it were.

Rather than impose such a highly intellectualized interpretation on this interactive episode, one could plausibly assume that the child, as children are wont to do, was picking up on the activity (game if you like) of pointing at objects and emitting sounds – different sounds for different objects. Roy does not give enough background to help us try to figure out what was actually going on in this situation; as is natural since he had a determinate notion of exactly what context was relevant for understanding the child's gesture.

Roy is far from alone in adopting this conception of what learning to speak consists in. Its philosophical interest lies exactly in its giving concrete expression to a line of thought which many of us (I for one) are inclined to embrace when thinking about learning to speak. When these matters are discussed, an episode in Helen Keller's autobiography is often cited, and Roy is no exception: the story of how she suddenly realizes that the tactile sign for water her teacher forms in her hand has something to do with the water gushing over her hands. If we scrutinize this anecdote more closely, we will note that it is hard to know how we are to understand it. What is actually involved in Helen's realization? Are we to think that she in a flash forms an idea of the entire human network of using signs to interact; does she see a connection between the repeated signings to which she has been subjected over the past (months?) and the words she had at an earlier time learned to speak; or is the whole episode a creation in the mind of her teacher, Anne Sullivan, who co-wrote the book? Unfortunately, well-known as this story is, I have never seen a critical discussion of it.

What makes an account in terms of reference seem so attractive is the fact that in the act of connecting a sound with an object we seem to have a concrete, almost bodily illustration of how words are bound up with reality, and thus, how meaning is able to seep into the learner's mind.

The idea that we learn to speak by learning to connect names with objects is of course also articulated in the famous passage from St Augustine's *Confessions* that Wittgenstein quotes at the beginning of his main work, *Philosophical Investigations*:

> When grown-ups named some object and at the same time turned towards it, I perceived this and I grasped that the thing was signified by the sound they uttered, since they meant to point *it* out. … In this way, little by little, I learnt to understand what things the words, which I heard uttered in their respective places in various sentences, signified. And once I got my tongue around these signs, I used them to express my wishes.[4]

Wittgenstein, who greatly admired Augustine, thought this passage was an eloquent expression of a conception of what it means to learn to talk which many of us will find natural. It constitutes the backdrop to Wittgenstein's introduction of the concept of a language game.

After quoting Augustine Wittgenstein goes on to imagine the so called builders' game. A is building something, B helps him by bringing him building stones when A calls for them: "block", "pillar", "slab", "beam". Wittgenstein asks us to "[c]onceive of this as a complete primitive language." (I am not quite sure of the spirit in which this exhortation is to be taken.)

Wittgenstein's aim here is to steer his reader away from the Augustinian conception. His strategy is what has sometimes been referred to as Wittgensteinian irony. In this case, it consists in going along with the account proposed by Augustine in order to see whether it will really yield the result it was meant to yield. For one thing, even if we imagine a language restricted to referring to objects, it is hard to imagine the distinction between common names and proper names being introduced through mere gestures of pointing? And what about number words, demonstratives? Even these simple names of objects require a context of human activity in which the relation between name and thing named is constituted. Furthermore, it is evident that a language consisting only of names of objects, as imagined by Augustine, is extremely impoverished. In fact, it is not even clear what this means. The point (or at least one point) is to make us realize how little is to be achieved through the gesture of pointing and vocalizing a word.

I will get back to the discussion of the builders later.

---

[4] Augustine, Confessions, I.8. Quoted in Wittgenstein, *Philosophical Investigations* (4th ed., Chichester: Wiley-Blackwell, 2009), § 1.

2. Observing communication

Literature offers another famous account of learning to speak. This is in Mary Shelley's novel *Frankenstein.* The monster, which Frankenstein has created out of body parts glued together and enlivened by administering an electric shock, is in some ways a precursor of our latter-day robots.

When the monster comes alive, Frankenstein runs away in shock at the sight of him and leaves him to fend for himself. After roaming about, the monster happens on a family living in a cottage in the forest. Having discovered that his looks tend to give people a fright he avoids making himself known to the family; however he hides in a woodshed from which without being observed he has a clear view of the family's life through a hole in the wall, rather as if he had been watching them on a television screen. The monster's story is reminiscent of Augustine's:

> By degrees I made a discovery of still greater moment. I found that these people had a method of communicating their experience and feelings to one another by articulate sounds. I perceived that the words they spoke sometimes produced pleasure or pain, smiles or sadness, in the minds and countenances of the hearers. This was indeed a godlike science, and I ardently desired to become acquainted with it. But I was baffled by every attempt I made for this purpose. Their pronunciation was quick; and the words they uttered, not having any apparent connexion with visible objects, I was unable to discover any clue by which I could unravel the mystery of their reference. By great application, however, … I discovered the names that were given to some of the most familiar objects of discourse; I learned and applied the words *fire, milk, bread,* and *wood.* I learned also the names of the cottagers themselves… I cannot describe the delight I felt when I learned the ideas appropriate to each of these sounds… I distinguished several other words, without being able as yet to understand or apply them; such as *good, dearest, unhappy.* (pp. 129 f.)

The mind of Frankenstein's monster, like that of Deb Roy's robot, is a paradigmatic Lockean *tabula rasa.* And as was the case for Roy and Augustine, the referential connection is here taken to be the clue to language. However, the monster's account adds an important dimension. Indeed, the starting point of his discovery and his motivation for wishing to learn this skill is that people use their voice to *communicate*, for instance, in producing pleasure or pain, smiles or sadness in their hearers. In other words, unlike what was the case with Roy, an idea of *what words are for* enters here. It is just that he, or rather Mary Shelley (who had been brought up in the tradition of British empiricism) takes it for granted that the basis of communication must be a shared system of reference. Anyway, at the end of this process, it turns out, the monster has learnt enough to be able to carry on an eloquent conversation the next time he meets his creator.

This story, of course, is no less fantastic than that of Augustine. For one thing, as in the case of Augustine, the monster is credited with the possession of fairly sophisticated ideas to start with. Thus, one might ask, what notion of *communication* might one have independently of language. We would not really expect a child to become a speaker in some such way; any more than, say, by shutting her up in a room and exposing her to endless hours of Youtube. What Mary Shelley is describing, again, is a case of learning without doing. It seems a mystery how someone could acquire a competence without ever having practised it, or, alternatively having had it explained to them (the latter of course is ruled out in the case of learning to speak). Frankenstein's monster, like Roy's robot and Augustine, are made out to be passive recipients of a conventional code, created in order to enable people and other beings to makes themselves understood to one another by referring to the objects surrounding them.

There is of course a famous critique of the empiricist notion of learning to speak, that primarily associated with Noam Chomsky.[5] Chomsky takes the bull by the horns and declares, in effect, that the new-born baby *must indeed* be in possession of the rudiments of language at birth. The problem with the *tabula rasa* is that it lacks the cognitive structure required to extract a grammar from the samples of speech with which the child is confronted. We can only assume that the child is equipped with a machinery for forming and testing hypotheses about the grammatical properties of the speech sounds she hears. Thus, daunted by the Scylla of the *tabula rasa* we might be tempted to address the Charybdis of innate grammar.

Chomsky is right, of course, in as far as a child must have a *readiness* for speaking of one kind or another. A chair will not start speaking, nor will a horse. For my present concerns, however, the difference between empiricism and nativism does not matter. No matter how complex the innate apparatus for learning to speak, the apparatus must gradually make contact with the outside world. The ability to form grammatically correct sentences is not enough. The question is how this can happen: under what conditions can a child grow into what we might call a linguistic agent: i.e. into someone who is capable of using words in order to express herself and make herself understood to others?

There is no reason why a *robot* could not be a Chomskian learner, i.e. have a built-in capacity to parse utterances and to construct syntactic hypotheses and test them out in practice. However, on *one* score a Chomskian robot would be in the same position as a *tabula rasa* robot. What appears to be lacking in either case is an account of how this robot comes to

5 See Chomsky, Review of Skinner's *Verbal Behavior*. Language. 1959; 35:26–58.

acquire the interactive habits which characterize speakers of a language – the habits manifested in our speaking to people and responding to their speech in meaningful ways.[6]

3. Becoming a speaker

As a child develops, she learns to call on the attention of others. At some stage words come into this. The child does not just express her discomfort into the blue: she addresses someone with sounds, gestures or words. If she fails to elicit a fitting response, she may repeat the signal, try something different or express her frustration. Gradually, too, she will learn to recognize that others are calling on *her* attention. Here also words enter at some point. In this way, she grows into the roles of speaker and listener. To become a speaker, we might say, is to become a participant in a network of roles. The reciprocity is essential.

In fact, to speak only of *a* role of speaker and of listener is misleading. The roles we acquire are a variety. There is requesting, demanding, reporting, exclaiming (out of joy, surprise, fear, anger, etc), asking questions, expressing affection, consoling, thanking, asking forgiveness, etc.[7] And their counterparts: complying, obeying, registering, answering, etc. The roles may often be intertwined. Thus becoming a speaker is not *one* thing but involves multiple forms of interaction; learning each is a separate process. These forms of interaction grow out of the child's relations with her elders. Her addressing others or taking herself to be addressed by them are not just random events. Their sense is bound up with the personal interconnection of which they are a part.

What I am adducing here are of course perfectly familiar facts. My purpose in reminding us of them is to turn our focus away from the sorts of concern that tend to preoccupy us when we address the issue of learning to speak in philosophical contexts. For one thing, learning to speak is not basically a cognitive venture, but a process of growing into an interactive network. The cognitive structures we invoke in characterizing people's linguistic abilities are not something that lies at the foundation of our interactive practices, but rather something we read out of (or should we say: read into?) the practices. The relevant cognitions, rather than linguistic ones ("The word 'Water' stands for such and such a substance", etc), are matters like, "There's someone there who might give me some water", etc. For another thing, the interesting development is not something to be found by searching the dark recesses of the individual mind, but something happening out in the open, in the spaces shared by individuals. We might say: the central question when we wonder how a child comes to be able to make herself understood is not: how is she *able* to say what she says,

[6] For a discussion of related issues, see Stuart Shanker, "What Children Know when they Know what a Name is" (*Current Anthropology* 42 (2001), pp. 481-497; especially pp. 486 f.
[7] For a sample of the various things people may do with words, see Wittgenstein, *Philosophical Investigations*, § 22.

but: *why* does she say what she says? What is she up to? What is she doing or trying to do? These are the questions that Roy apparently failed to ask in the course of his inquiry.

4. Routines and skills

Let us get back to the builders' game. It illustrates two simple kinds of interactive roles:

(1) The helper's role. He has to bring the stone called for by the builder. In this sense, what he has to do is fixed: within the confines of the game as described by Wittgenstein he has no alternative options. If he does not bring the stone, or if he does something different, he is no longer playing this game. To play it he must be able to recognize different kinds of stone, but since the task of recognition is a simple one, and the requisite action is straightforward, we could imagine his learning his role entirely by rote.

Reciprocally, many *uses* of words are learnt by rote: say, simple games like "This little pig went to market", saying the alphabet, counting objects, doing the multiplication tables. Let us call them, as well as the required responses, verbal routines. Such performances are of course easy to simulate on a robot. It might be programmed to say (in the right circumstances): "My batteries need recharging", "There's a leak in the nuclear plant", or "We are arriving at Brisbane Central Station"; equally it might be programmed to respond to voice commands. Anyone who knows English would understand the robot's messages or be able to give commands without needing to know anything about robots. One would simply trust the programming.

There are other uses of words that are routine-like, yet involve a higher degree of complexity than these, say greeting or thanking someone. We might call them semi-routines. Thus, in certain fairly well-defined situations saying "Thank you", or closely similar words, is the only acceptable response. Yet recognizing those situations requires a degree of cultural competence. You would not thank me if I handed you the hammer to hold while I put a mark on the wall, but you would if I handed it to you in order to help you, or as a gift. Could a robot learn to recognize these situations? Could there be such a thing as a gift to a robot – can a robot *own* things? Is it even clear what "helping" a robot might be? (Do I help my car when I push-start it?) Sometimes the help may be quite subtle, such as supplying the word I was looking for. But do robots ever look for words? I short: does a robot ever have anything to be thankful for?

(2) The builder. He is bound to a limited vocabulary – the names of building-stones – and a limited range of verbal activity – telling the helper to bring him a stone. Yet he has a range of options when it comes to choosing which stone to ask for. His use of words constitutes a low-level *skill* – he is able to use words to bring about certain specified results. The words he uses are mediators between the material circumstances and the end-result. There is no need

for him to adjust his words to an individual helper: any helper who has been taught his task will respond in the same way.

In the simplest cases, the counterpart of a verbal skill is a routine, as in the case of the builder's helper. There should be no obstacle in principle to a robot being programmed to exercise certain kinds of verbal skills and respond to their exercise.

5. There's more to speaking

Verbal routines (and semi-routines), then, are a matter of conforming to a pattern in speaking or responding. Verbal skills, in turn, are a matter of bringing about a specific result, by addressing someone or after being addressed by someone. However, there are many dimensions of human verbal interaction that cannot be captured in these terms. (I should point out that routines and skills are not sharply demarcated categories but that aspects of them may be present to a higher or lower degree in all our speaking.)

Consider, for instance, verbal reactive expressions of pain, disgust, joy, surprise, pleasure, anger, fear, etc. As Wittgenstein suggests, we may think of them as replacements of the natural expressions of those sensations and feelings:

> Here is one possibility: words are connected with the primitive, natural, expressions of sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour.[8]

These expressions may simply be wrung from me, or they may constitute a plea for help or consolation, a shameful confession, etc. Could a robot, say, moan with pain? Of course it could be programmed to simulate moaning, in the sense that it could be caused to produce sounds that resemble, or may even be mistaken for, human moaning. But the sound by itself does not make something a case of moaning with pain. It must be a *genuine* moan. We may be tempted here to say that the moaning must be bound up with an *inner* occurrence. However, as Wittgenstein makes clear, if by something "inner" we mean a process to which no one else may have access, this "inner process" cannot be appealed to as a test for the authenticity of the expression.[9] To experience the expression as genuine *is to experience it as an expression of pain*, and *vice versa.* One person may credit the expression, another may not (he may have reason to suspect that I am simulating pain). There are no independent criteria by which it is to be decided who is right, although closer observation or closer

[8] *Philosophical Investigations*, § 244.

[9] On this, cp Wittgenstein's comparison of the inner to a beetle in a box, *Philosophical Investigations*, § 293.

acquaintance with me or a different understanding of the situation may bring one of the observers around.

How do spontaneous verbal expressions come about? Consider, say, a child learning to ask for a drink of water. This is hardly to be understood in terms of the child's recognizing that she is thirsty, then developing a technique for setting in motion a process that will ultimately lead to her having her thirst quenched; rather in learning to ask for a drink the child develops an understanding of what it means to be thirsty. This is *part* of the story: there is of course an element of reciprocity in learning to understand about thirst: I do not know what it means to be thirsty unless I realize (whether I act on it or not) that someone else's expression of thirst may involve a call on me to give him something to drink. Asking for a drink is neither a routine nor a skill. Learning to express thirst or learning to understand expressions of thirst does not consist in acquiring a technique for achieving certain ends. A more natural way of describing what goes on is to say that the word "thirst" comes to be incorporated into the child's life, into her relations to the people around her. There is no meaning kernel that can be considered in separation from the forms of social interaction in which the word "thirst" has its use. How can we tell that a child has mastered this use of words? We see it from the circumstances in which she uses them and her behaviour – whether she actually drinks the water we give her, for instance. Of course, a child may ask for a drink just to put off having to go to sleep, but that tactic will only work once her elders have come to believe that she has mastered this use of words (and it will stop working once they see through her ruse).

We are beings to whom sensations like thirst, or wishes, intentions, etc can be attributed because we are beings who may acquire the ability to say we are thirsty, we wish for something, intend to do something, etc. (It is tempting to say we have wishes because we can say we have wishes, but that would be a simplification.) Wittgenstein writes:

> By nature and by a particular training, a particular education, we are predisposed to express wishes in certain circumstances. (A wish is, of course, not such a 'circumstance'.) (PI § 441.)

6. Expressive robots?

For something to constitute for me a genuine expression of pain, joy, fear, etc is connected with its being bound up with *the life* of a being, whether human or animal. A robot, on the other hand, does not have a life. Suppose, on the other hand, someone were to insist that a robot does have a life, *what* is it that he would be attributing the life to? The more or less humanoid or animal-like entity that is moving before us and where sounds are coming from, or the software which is perhaps somewhere else, which records the inputs of the robot and which guides its functioning? The robot in front of me, we might say, is the physical

instantiation of a range of programmes and a data base, which may of course be equally instantiated in innumerable other robots. So it is not a particular speaker. As a speaker it might be compared to a ventriloquist's doll: the source of the speech is not where the sounds are produced or seem to be produced.

As long as I keep in mind that what I have before me is a robot (no matter how humanoid its external design), it will be clear to me that the moaning is pre-programmed to occur when it does. This means that there is no room for raising the question whether the robot's moaning is genuine of fake. In the robot's "behavioural repertoire" (if we may call it such) there is no room for anything to which the grammatical categories – the genuine/fake distinction - of expressions of pain, or other expressions of feeling or sensation, applies. And the same goes for the verbal counterpart of those expressions. – "But suppose the moaning-like sounds are produced by some malfunctioning of the hardware – say, that the fans need lubricating." Still, unlike human expressions of pain, this would not be a manifestation of an "ability" on the part of the robot, any more than the creaking of a door displays an ability on the part of the door. For a human being to be moaning with pain may be caused, if you like, by a malfunctioning of the body, but the moaning itself is not a malfunction.

This is not to say that we may not respond to robotic expressions *as though* they were real. In care for the elderly, for instance, robots are sometimes used to provide a sense of companionship, to forestall loneliness and dejection. One of the best-known health care robots is the Japanese robotic seal Paro. Even though a client is clear that what he has before him is a piece of machinery in animal form, he may still feel some sense of encouragement from the seal's responses. This could be compared to a conjurer's trick: we may have a strong sense that the magician is picking eggs out of his assistant's ear; even so, most of us will not really believe that that is where the eggs came from. We simply let ourselves be lulled into a form of oblivion, we lean back and enjoy the paradox.

A nursing home in Pittsburgh employed a Paro to provide companionship for the residents. Dr Marette, a clinical psychologist working in the facility

> said she initially presumed that those who responded to Paro did not realize it was a robot — or that they forgot it between visits.
>
> Yet several patients whose mental faculties are entirely intact have made special visits to her office to see the robotic harp seal.
>
> "I know that this isn't an animal," said Pierre Carter, 62, smiling down at the robot he calls Fluffy. "But it brings out natural feelings."
>
> Then Dr. Marette acknowledged an observation she had made of her own behavior: "It's hard to walk down the hall with it cooing and making noises and not start talking to it. I had a car that I used to talk to that was a lot less responsive."

> That effect, computer science experts said, stems from what appears to be a basic human reflex to treat objects that respond to their surroundings as alive, even when we know perfectly well that they are not.[10]

We should note that while Dr Marette and Mr Carter speak about reacting to Paro as though it were alive, they are never tempted to claim that the robot may indeed be expressing actual feelings.

7. Spontaneity in speaking

Rush Rhees, who was student of Wittgenstein's and who devoted much of his work in philosophy to the exposition and discussion of Wittgenstein's work (besides being one of the editors of Wittgenstein's written *Nachlaß*) was quite critical of Wittgenstein's use of the builder's game; more specifically, of his exhorting us to think of the game as a complete language. This was the theme of his essay "Wittgenstein's Builders".[11]

There were two points in particular that Rhees wanted to emphasize. First: language forms a unity. Not a formal unity in the sense of deductive logic, rather a unity in the sense that different occasions for speaking hang together, bear on one another:

> In learning to speak [the child] learns what can be said; he learns, however fumblingly, what it makes sense to say. He comes to have some sense of how different remarks have something to do with one another. … in beginning to carry on a conversation, in trying to tell you something and trying to understand your answer, he *is* getting a sense of how different remarks have a bearing on one another. And because he learns this, he can go on speaking or go on learning. (Pp.79 f.)

Second: in her becoming a speaker, words become a way for the child to express herself. She comes to use words spontaneously, not just out of routine or because the situation demands a particular response. Words become part of who she is. They are not something external, imposed on her. Learning to speak is not

> like learning a game. We may *use* something like a game in teaching [the child]. We say his sounds back to him, and in this way we bring him to imitate other

---

[10] Amy Harmon, "A Soft Spot for Circuitry", *The New York Times,* July 4, 2010. https://www.nytimes.com/2010/07/05/science/05robot.html?_r=2&pagewanted=1

[11] Originally in *Proceedings of the Aristotelian Society*, 1959-60. Reprinted in Rhees, *Discussions of Wittgenstein* (London: Routledge and Kegan Paul, 1970). I quote from there. -- A slightly amended version occurred in Rhees, *Wittgenstein and the Possibility of Discourse*, ed. D. Z. Phillips (Cambridge University Press, 1998).

> sounds we make. And this is a game. But it is not what we are trying to teach him. And if all he learns is how to play like this, he will not have learned to speak. He will never tell you anything nor ask you anything either.
>
> When he can speak we may be delighted because 'He can say things himself now – not just repeat'. But what is important is that he can *say* things: not that he can construct new sentences – as it were in an exercise. (P.80.)
>
> If someone learns to speak, he does not just learn to make sentences and utter them. Nor can he merely have learned to react to orders. If that were all he ever did, I should not imagine that he could speak, and I should never ask him anything. When he learns to speak, he learns to tell you something; and he tries to. (P. 79.)[12]

These two points – the unity of language, and spontaneity – hang together. It is because the uses of words are not tied to specific situations that we are able to use them to express ourselves.

The dimension of spontaneity characterizes a very important part of the lives we live together, the ways we speak. For someone's use of words to be spontaneous, in the sense I am talking about, she need not be particularly lively or quick on her feet. I mean simply that she says things even though they are not required by convention or by some urgent practical need, yet not speaking randomly, but rather saying things that it makes sense for her to say at the time of speaking. A married couple over breakfast may exchange ritual expressions (how did you sleep?), they may use verbal skills (could you pass me the butter? or, on a more advanced level: will you need the car today? etc), or express themselves (oh, I'm so tired!) But if they are never spontaneous in what they say, you might be worried about their relationship. That is, suppose they never volunteer things like: how did you like the movie we saw last night? or: did you read the latest news about the Brexit vote in parliament? or: where should we go for our holiday this year? In a certain sense, they would not be speaking to one another. They would not be related to one another as speakers.

To be a speaker is not simply to have mastered certain verbal routines or verbal skills. I believe that is what Rhees was emphasizing. But how would we go about training a child to speak spontaneously? That is the problem: for to speak spontaneously is precisely to speak independently of any specific requirements, and yet to speak in a way that is intelligible in the conversational context. The speaker should express herself, and in doing so make

[12] Arthur Coady puts a similar point as follows in his thought-provoking essay "On the Creation of a Speaker": "For me to speak there must be an understanding of what there is to say. That understanding must be *mine.*" *Mind* 77 (1968), pp. 68-76. The quotation is from p. 75

contact with those she is addressing. It seems there are no routines or drills we can go through in order to teach a child to speak spontaneously. The child will just have to latch on to speaking. There is not some separate thing that needs to be done in order to reach the stage of spontaneous conversation. We learn conversation in taking part in conversation.

This could also be expressed by saying that the child will not have a grasp of herself as learning something, of there being something she needs to learn. This seems to be what lies behind the point made by Coady, in "On the Creation of a Speaker" :

> A child puts forth no effort to learn his native tongue (ignore learning grammar or big words), but he does struggle. Bewilderment, confusion, failure must be overcome, though it takes no effort. (P. 74.)

I take it that, for Coady, an effort entails striving towards a goal of which one is conscious. A child is not conscious of speaking as a goal she is trying to attain – unlike the case of someone learning a second language. Yet in a particular situation she may struggle to "get it right", that is, to put things in a way others will understand or accept.

8. Spontaneity in robots?

How, then, could one programme spontaneity into a robot? A robot has no life that it could learn to express. Perhaps one might programme it to make certain utterances at the occurrence of certain inner states. But that is not what we mean by spontaneity. Saying "I'm thirsty" is not a response to a state of dehydration. It is a way of verbalising a wish. The robot does not have wishes in this sense. Alternatively, the robot could be programmed to "bring up" various topics at random points in time. But neither is that what we mean by spontaneity. To use words to express oneself is not to speak at random.

The things we say may be a reflection of our temperament, our mood and the mood of the situation. They are – or we take them to be – suited to the person we are addressing, to our relation to her, to the wider context, to the history we share, or do not share. The role one assumes in a conversation will be an expression of one's life. A life, however, is precisely what a robot lacks.

A human speaker's relation to his words has depth:

> … if I know you can speak, then it makes sense for me to ask you what you mean, to try to get you to say more clearly what you want, and to ask you

> questions about it … The example of the builders does not seem to allow for any of these.[13] (P. 81.)

With a human speaker, the question of whether she really means what she says may arise. We may on occasion have to search our soul in order to be clear whether we really meant what we were saying. Did I only say that to be polite? was I carried way by the general mood? was I putting on an act? did I speak without thinking? did my emotions get the best of me? did he cajole or frighten me out of my determination? etc. The fact that questions like these may be raised is part of what it means to be a speaker. There is no room for such worries where the speech of robots is concerned.

Maybe this was what Deb Roy realized, and why he gave up the attempt to teach his robot to speak by the method he had used to teach his son. Or better yet: maybe he realized that his son did not become a speaker because of any *methods* he used.[14]

[13] Rhees, *op.cit.,* p. 81.

[14] I wish to thank Merete Mazzarella, as well as the participants in the philosophy research seminar at the University of Queensland for helpful comments.