AUTONOMOUS ROBOTS AND TACIT KNOWLEDGE

Mihály Héder and Daniel Paksi

Abstract
The recent decade has seen so much development in the area of autonomous robots that it is worth (re-) investigating the relation of machines to knowledge according to the concept of tacit knowledge put forth by Michael Polanyi. In this paper we argue that certain machines—autonomous robots with a 'centre'—have tacit knowledge, much in the same way that animals do. They cannot explicate their knowledge, and their knowledge is not identical with the partly explicit knowledge that an engineer possesses about the machine, e.g. its program code. When someone tries to understand how a robot operates, it is very easy to unreflectively mix these points of view and come to the false conclusion that robots are capable of doing things because they are explicitly instructed to do so.

Key Words
Michael Polanyi, tacit knowledge, autonomous, machines, centre, boundary conditions

1. Introduction

The recent technology exposition iRex 2011 included a great attraction for the visitors: a robot called Primer-V2 riding a tiny bicycle (see the video: www.youtube.com/watch?v=mT3vfSQePcs) This robot is a humanoid, meaning that its 40cm tall body, made of aluminium and plastic, mimics the human form. The robot has four different sensors that provide feedback to the central control unit, which is located in the backpack of the robot and programmed on a chip slightly larger than 1x1 cm.

A remote control was used to direct the robot, but it only sent high-level commands like 'bike forward' or 'stop'. Pedalling and balancing are managed by the robot itself. The Japanese creator's next goal is to enhance the robot to allow it to plan its own route, thus making the remote control unnecessary.

The Primer-V2 represents only the latest stage in the evolution of bicycle-riding robots. It is especially interesting to us because of its autonomy, its ability to balance without a gyroscope, and its humanoid body. Even the bicycle is a regular one, only a little bit smaller.

Many of these features were present in earlier projects as well, but not at the same time. The humanoid Murata Boy already pedalled on a bike in 2005; its sister, Murata Girl was able to ride a unicycle. These robots were stabilized by a gyroscope. Other robots, like the entrants in the BicyRobo Thailand student challenge (first organized in 2011) did not use additional stabilizators, balancing using handlebars only. However, these robots did not have a rider, but were automated bikes.

(*See: www.filozofia.bme.hu/pub/appraisal/robots/Figure2.jpg*)

There are many other examples through which we could investigate the question of machine knowledge: e.g. chess-playing robots, the Wolfram Alpha question answering system, the famous jeopardy player Watson, unmanned aerial vehicles (UAVs), the Mars Rovers, etc. But we are sure that for Polanyi readers it is clear why we choose this particular robot: bicycle riding is one of Polanyi's favourite examples for explaining tacit knowledge (e.g. 'The Logic of Tacit Inference', KB, p. 138-158,[1] or PK pp. 49-50).

Our goal in this paper is to evaluate the relation of tacit knowledge and autonomous robots like Primer-V2. These robots could only be speculative thought experiments in Polanyi's time. This means that we are applying the concept of tacit knowledge in a new area; we think this application is possible without changing or abandoning the original philosophical framework. Primer-V2 appears as something that rides a bike. Is this achieved with knowledge of same kind that humans have? Is it fundamentally different? Or should we say that it is not knowledge at all—only successful operation?

2. Emergent organisms and machines

The first question we have to answer concerns the ontological status of machines. Polanyi discusses machines and living organisms together in 'Life's Irreducible Structure' (Polanyi, 1968a). Is this just an analogy or more?

Machines are emergent entities as they are controlled by two different principles irreducible to each other:

> A machine as a whole works under the control of two distinct principles. The higher one is the principle of the machine's design, and this harnesses the lower one, which consists in the physical-chemical processes on which the machine relies (p.1).

This structure is the basis of living organisms as well: 'Living Mechanisms are classed with machines' (p.1).

By establishing that machines and living things belong to the same class, a straightforward means of explaining certain features of the living opens up:

> Morphogenesis, the process by which the structure of living beings develops, can then be likened to the shaping of a machine which will act as a boundary for the laws of inanimate nature. For just as these laws serve the machine, so they serve also the developed organism (p.1).

A consequence of the emergent, two-layer structure of machines is that (like life) they cannot be explained on the lower, explicit physical or chemical level alone:

> Engineering and physics are two different sciences. Engineering includes the operational principles of machines and some knowledge of physics bearing on these principles. Physics and chemistry, on the other hand, include no knowledge of the operational principles of machines. Hence a complete physical and chemical topography of an object would not tell us whether it is a machine, and if so, how it works, and for what purpose. Physical and chemical investigations of a machine are meaningless, unless undertaken with a bearing on the previously established operational principles of the machine (TD, p. 39).

We have to come to the conclusion that in Polanyi's view machines and living organisms belong to two *different subclasses* of the *same class of emergent entities* under dual control.

It is also clear that living organisms have tacit knowledge. As Polanyi states in Personal Knowledge:

> *'knowing belongs to the class of achievements that are comprised by all forms of living,* simply because every manifestation of life is a technical achievement (...) (PK, p. 403).

The second part of the sentence is especially interesting for us as, because it makes clear that the classification we explained above is in fact *ontological,* and also that knowing is a feature of this class *in general.* It is also a denial of the materialist view that life's unique phenomena are a result of a random coincidence of physical and chemical processes.

Polanyi makes his position even more clear when he discusses machines and simple organisms such as the amoeba:

> I think that what you call the logic of choice is deeply imbedded in all manifestations of rationality down to the level of the amoeba It is likewise inherent in the conception of all machines and indeed of any purposive device (Polanyi, 1953).

Or even simpler life forms (PK, p. 387 'bacillus', p. 400-401 'germs').

As we already pointed out, the structures of machines and living things are more than similar. Polanyi distinguishes *two* different types of boundary conditions ('Life's Irreducible Structure'). The first is the *test-tube* type boundary condition, which ontologically does not transcend the level of physical-chemical processes. Contrary to this, the *machine* type boundary condition *transcends* the lower level. 'Thus the morphology of living things transcends the laws of physics and chemistry' (p.2). As in the case of machines, where the principles of engineering are needed in addition to physical-chemical principles to control the machine's physical structure and achieve its goal, living things have their own biological principles in addition to physical-chemical ones. In other words, according to Polanyi, because of the machine type boundary conditions, machines and living organisms are different from other entities like a crystal or a tornado that fall under only test-tube-like boundary conditions, or in other words that are only governed by physical-chemical principles. This reassures us that machines and living organisms belong to two subclasses of the same class.[2]

Polanyi explains how important the 'unformalizable regulative functions'—which belong to the higher level of the emergent structure—are for supporting life. We can find similar regulative functions in the case of Primer-V2: the control unit in the backpack of the robot provides these functions.

(P1) Considering that machines and living things are subclasses of the same ontological class, and (P2) recognizing the control mechanism of a robot as the machine equivalent of living organism's regulative functions, while not forgetting that (P3) all forms of life are capable of knowing, that is, have some kind of knowledge, we arrive to the conclusion (C) that robots like Primer-V2 also possess some kind of knowledge.

It is important to point out that Primer-V2's machine-like, emergent structure alone is *not enough* to explain that it can possess knowledge. An additional requirement is needed to fulfil this ability: that it has a *centre* (PK p. 344) that features regulative functions that control its body and maintain its operation, '…a centre of self-interest against the world-wide drift of meaningless happenings' (PK p. 387).

The concept of the regulative centre enables us to resolve the deep problem generated by the fact that robots are not living organisms and yet they know certain things. The presence of a centre is *necessary* for even the most primitive forms of life, because they would not survive a minute without the regulative functions realized therein. In the case of machines however, a centre is not necessary.

Humanity has invented many machines—like the hammer, or the bicycle—that fall under the dual control of test-tube type and machine type conditions, but do not have a centre; this category does not exist in the case of biological life. These machines are not autonomous and require an operator. We can see these machines as *extensions* of the human body, as in Polanyi's example of a man who orientates himself with a staff. Or we can say that these machines are regulated by their operator's centre. In other words, while Primer-V2 has its own knowledge, a hammer does not—the man with a hammer does.

In this article we only discuss machines that are autonomous, a feature achieved by a regulative centre. We do not say that a hammer knows how to nail, or a car knows how to accelerate, etc., we discuss autonomous robots only. Nevertheless, it is a non-trivial task to define the boundary between autonomous robots that have centres and simple machines and tools that do not. We think that according to Polanyi there is no clear, explicit definition for this boundary. In any case, we are not aiming to answer this question here, but are asserting that robots like Primer-V2 or an autonomous UAV definitely fall in the category of machines with centres and tools like a hammer or a (regular) bicycle fall definitely outside of it.

It is very difficult to deny the capacity for any kind of knowledge in the case of Primer-V2 or similar robots. In this position one has to argue that the robot does not know how to ride a bicycle, even though it does something very similar; or that a chess machine does not know how to play chess, even though a layman cannot beat it anymore at the game.

3. Designing a knowing robot

In the previous section we explained that in certain cases we have to call a robot's performance 'knowledge'. Now it is time to discuss how Primer-V2's knowledge of riding a bicycle relates to similar knowledge employed by a living organism, in this case a human or a trained primate.

It is clear that Primer-V2's capacity for bicycle riding is achieved in a very *different* manner from the way a human achieves the same. In the robot's case the regulative functions are realized with a proportional-integral-derivative (PID) method, a classical approach in control theory; this regulation is very different from what a human or an animal does. It is also evident that the body structure of the robot is very different from the human body: its stability is not provided by a skeleton, there are no muscles, and the motion is achieved by servo motors, etc. (*see http://www.filozofia.bme.hu/pub/appraisal/robots/Figure3.jpg*), which displays the components of the Kondo HRV, a commercially available robot kit on which Primer-V2 is based).

Moreover, human bicycle riding was not fully explicated—something that is impossible anyway according to Polanyi—and therefore we cannot say that human knowledge is somehow being explicitly simulated by a robot.

To understand the situation better, let us consider Polanyi's example of the neurologist (Polanyi, 1968b, p. 39).

The neurologist is able to examine the brain of another person while that person is, for example, watching a cat. The scientist is able to make focal the subject's brain's internal processes. Of course the subject itself cannot do this.

> But the facts remains that to see a cat differs sharply from the knowledge of the mechanism of seeing a cat. They are a knowledge of quite different things (Polanyi, 1968b, p 39).

In other words no matter how fully the neurologist explicates the subject's brain mechanisms, the knowledge he gathers is *not the same* as the subject's own knowledge. As a consequence, the scientist can*not* use the subject's knowledge as his own. The deep meaning of this example is more evident if we consider the case of riding a bicycle or playing a piano. The neurologist might be able to give an exhaustive explicit description of how the subject rides the bicycle or plays the piano in terms of brain and body mechanisms, but of course having only this knowledge does not enable him to ride or play at all.

The engineer is in a similar situation to that of the neurologist, in that she understands the software and hardware required to build a bicycle-riding robot. One could argue that the engineer explicitly knows every instruction required to make Primer-V2 ride and therefore its knowledge is fully explicated; and one could then arrive at the conclusion that because fully explicated knowledge is impossible according to Polanyi ('The Logic of Tacit Inference', KB 138-158. p.144, see the quotation later in this section), in this case there is no knowledge at all.

However, *both steps are wrong*. The robot's program code or hardware blueprint as grasped by the engineer is like the explicated brain mechanisms understood by the neurologist. It can be made focal, it is discoverable, it might be even formalizable—but *it is not the robot's knowledge*. It is the knowledge *of a spectator about* the robot.

On Figure 1 (see below, p. 14) we explain how we interpret the general process of the construction of a bicycle riding robot.

1. A person rides a bicycle.

2. A scientist examines the human's bicycle riding skill.
3. The scientist explicates his knowledge about the subject in mathematical formulae. This is not the same as the subject's knowledge. As a corollary, the scientist is not able to acquire the subject's knowledge of riding—maybe the scientist does not know how to ride a bicycle at all; it is beside the point. This knowledge is similar to that of the neurologist about the brain.[3]
4. The scientist transforms his explicated knowledge about bicycle riding to hardware architecture and program code.
5. The scientist builds and programs the robot.
6. The robot rides the bicycle.

We have to point out that the knowledge of the actual, built robot is not explicit, even though the scientist programmed *his* explicit knowledge in it. First of all, this is because the robot has a body of aluminium and plastic and servo motors, etc., about which the scientist himself has no fully explicit knowledge. Second, the program has *different meanings* for the robot and the scientist. For the scientist it is explicit knowledge, a description of a control method. For the robot, it is not explicit knowledge about something, it is something it *applies*. The robot does not have the explicit knowledge of the scientist, it *does not even know the program*—it does not understand programming patterns, PID control, etc.—it knows how to *run* a program. In this case, by controlling its body structure according to the program, it knows how to ride a bicycle.

In other words, what the scientist explicitly expresses in the program code is not explicit *for* the robot. It does not know what is written in the programming language (the scientist's explicit knowledge) and it does not ride by understanding it. What it does is *execute* a code, *integrating* it with its body structure into physical motion, enacting the knowledge of bicycle riding itself (a *tacit* knowledge). The scientist, in general, cannot do such things. On the other hand, the robot does not have his explicit knowledge at all!

Of course one could program a robot to print its program code at the push of a button—but the printed material also will not be the robot's knowledge. One could even engineer a robot in a way that it would display its hardware blueprints—analogous to the situation of the medical student who writes down the entire anatomy atlas (PK p. 89). That would be the robot 'explicating' a part of the scientist's knowledge (which have tacit components in the scientist himself!), not the robot's own knowledge, which is restricted to the capability of running the code and thus riding the bike. As Polanyi explains:

> While tacit knowledge can be possessed by itself, explicit knowledge must rely on being tacitly understood and applied. Hence all knowledge is either *tacit* or *rooted in tacit knowledge*. A *wholly* explicit knowledge is unthinkable' ('The Logic of Tacit Inference', KB 138-158. p. 144).

We have to come to the conclusion that, just like with animals or humans, the robot's knowledge is at least partly—but more likely totally—tacit. We cannot say that the robot works according the explicit knowledge of its creator; we also cannot say that the tacit part of their knowledge is similar, as they work according to very different principles. The corollary of this proposition is that, although they both know, there is a major difference between the kinds of knowledge possessed by robots and animals or robots and humans.

This very important distinction between the human's and machine's tacit knowledge is essential for the philosophy of AI debates. Without this distinction a deep tension arises because identifying a human's knowledge with its brain processes, which scientists might exhaustively describe one day, while at the same time identifying the robot's knowledge with its program code causes the difference between the two to appear to vanish.

4. The problem of consciousness

The arguments about robots' knowledge generate strong feelings and vigorous denial in many audiences. We believe that, at the core of these feelings, many people think that if robots possessed knowledge, then they would be like us. However, robots are clearly not like us, and consequently they cannot have knowledge; this is how the reasoning continues. For illuminating the difference between robots and humans, a common argument is that while humans are conscious, robots are not. Moreover, this argument is supported by Polanyi himself:: he states that any kind of awareness, including the capability to adapt as well as knowledge, requires some degree of consciousness (Polanyi, PK, p. 92).

We want to emphasize that when arguing that robots have tacit knowledge, we do not mean, at the same time, that they are like humans. We should not forget that the kind of knowledge we attribute primarily to humans is explicit knowledge, which separates them from the animal kingdom. According to our argument, robots do not have explicit knowledge, and this means they are very different from humans. Moreover, as in tacit knowledge, the embodiment is crucial, and as robots have bodies different from ours, so their tacit knowledge is also fundamentally different from tacit knowledge in

humans or animals. Robots are really unlike us. Our argument extends only to stating that they possess knowledge, and that knowledge is tacit. (Very interestingly, the idea of a robot having explicit knowledge seems more acceptable to many. Actually, this idea is much more radical than robots possessing tacit knowledge, and it can really dissolve the distinction between humans and machines.)

Naturally, it is possible to draw a sharp line between machines and humans in respect to consciousness. In order to do that, one need only consider the classical, critical notion of consciousness that is rooted in Cartesian philosophy. According to this, consciousness is a transparent, purely rational, reflective phenomenon that is a feature of humans only. In this sense, neither animals nor robots have consciousness. We think that the argument for a lack of consciousness in the case of machines is a result of this classical view. In Polanyi's philosophy, we can talk about only this kind of reflexive consciousness at the level where there is language and explicit knowledge. This kind of consciousness enables the recognition and articulation of otherwise fully tacit thinking processes that in this way can become subjects of focal awareness. In this sense, following Polanyi's philosophy, the distinction is clear between robots and humans. However, from this narrower definition of consciousness, animals are also excluded. If this kind of consciousness is necessary for knowledge, then animals cannot have knowledge.

But in Polanyi's philosophy it is clear that reflective, explicit Cartesian consciousness is not a prerequisite of knowledge and that animals have tacit (and only tacit) knowledge too, as we have seen in Chapter 2:

> . . . knowing belongs to the class of achievements that are comprised by all forms of living, simply because every manifestation of life is a technical achievement, and is therefore—like the practice of technology—an applied knowledge of nature. (PK, p. 403)

And tacit knowledge involves a certain degree of consciousness:

> While focal awareness is necessarily conscious, subsidiary awareness may vary over all degrees of consciousness. (PK, p. 92)

Following the second quotation, Polanyi discusses animal consciousness and the active centre that is a prerequisite to it. This active centre is necessarily present in every living organism, but in the domain of machines, it is present only in autonomous robots.

As we can see, in Polanyi's emergent worldview, both consciousness and knowledge are gradually emerging, dynamic phenomena that are present in even the simplest living organisms. Autonomous robots are not living, but they are still emergent entities at a similar level to simple life forms. An amoeba is able to perceive the presence of food through its chemical receptors, to extend its body in the direction of the food, and finally to ingest the food. Likewise, a Mars Rover is able to perceive obstacles that are in its way through its visual receptors, and then it can calculate a new path and pass by the obstacle. At a very low level of consciousness, both the amoeba and a Mars Rover are aware of the food or the obstacle and each reacts accordingly. However, it would be really interesting research to investigate the degrees of consciousness in emergent development.

In a materialistic view—one that is heavily criticised and rejected by Polanyi—machines are purely material things, so their nature is not different from that of a rock or a cloud, and, by corollary, they cannot have knowledge. However, this argument does not stop with machines; it will necessarily extend to animals and humans as well. If we do not draw a line between purely physical objects and emergent beings—and in Polanyi's view machines are clearly emergent—with different levels of consciousness and knowledge, then how will it be possible to show that humans or other beings have more knowledge than a rock? On this path, we eliminate human thinking as well by reducing it to ion streams and electron transmissions.

5. Consequences and conclusions

The evaluation of machine intelligence is a difficult task that generates extensive debates. But the problem cannot be ignored because the systems of the 21st century continue to produce surprising results. For the sake of simplicity, in this article we only discussed a bicycle-riding robot. However, there are many other areas of interest: planetary explorer robots (Mars Rovers); autonomous ground and aerial vehicles usually utilized in combat; autonomous household robots; autonomous factories; autonomous life support systems in hospitals, etc. We think that our conclusions summarized in the following points hold for many applications:

(1) These robots, like all machines, are emergent, in other words they are not purely physical in essence. This is made clear by Polanyi himself in 'Life's Irreducible Structure'.

(2) Robots are capable of possessing knowledge; for instance, Primer-V2 knows how to ride a bicycle. However, this knowledge is fundamentally different from animal or human knowledge.

(3) A robot's knowledge is always at least partly tacit. This is absolutely consistent with the way Polanyi uses the term. In our case with the

Primer-V2, Polanyi's explanation of how bicycle riding must be based on tacit knowledge is still valid, as the robot's knowledge relies on the tacit integration of an aluminium and plastic body and a program code that is explicit only for the programmer.

We do not think that points (2) or (3) ever occurred to Polanyi himself, or that he even considered this problem in a similar manner as we have. However, he did consider the problem of the Turing test and the simulation of mind (PK p. 263):

> Mind is not the aggregate of its focally known manifestations, but is that on which we focus our attention while being subsidiarily aware of its manifestations. […] According to these definitions of 'mind' or 'person', neither a machine, nor a neurological model, nor an equivalent robot, can be said to think, feel, imagine, desire, or judge something. They may conceivably simulate these propensities to such an extent as to deceive us altogether [Polanyi refers here to the Turing test]. But a deception, however compelling, does not qualify thereby as truth: no amount of subsequent experience can justify us in accepting as identical two things known from the start to be different in their nature.

These views on the simulation of the human mind support our conclusions. Polanyi, although his discussion of the problem is brief, instantly recognizes that it is *impossible* to identify a human mind with a machine that 'simulates' what is explicitly known about it. This is because of to the simple fact that the mind itself is *more* than its explicit description. In this article we rely on this fact and go one step further to state that *it is also impossible to identify a machine with its explicit description*. From this it is clear that a machine which is equal to humans is a logical impossibility. Polanyi's notes on this matter have only recently been published: Polanyi, 2010, p. 97. Polanyi Archives: Michael Polanyi on Mind and Machine).

Other variations of Polanyi's arguments concern the ineliminable human element in deduction carried out by machines (PK, p. 257-258) and the McCullogh-Pitts neural model's inability to explain intelligent behaviour (PK, p. 340). However, as we have tried to point out, his arguments against any possible equation between a human mind and machine are not inconsistent with a machine possessing tacit knowledge.

We think that our arrival at conclusions that Polanyi himself never intended to draw only emphasizes the truth of his philosophy: '… truth lies in the achievement of a contact with reality—a contact destined to reveal itself further by an indefinite range of yet unforeseen consequences' (PK, p. 147).

Following up on our analysis, there are other possible questions to investigate in the future. One could investigate networked robots or a broad range of computers and their relation to knowledge. A proper definition of a centre in robots would be needed—probably based on systems theory. An even harder question is that of articulation made by robots. For Polanyi, the ability of articulation distinguishes humans from other living things. What if we come to the conclusion that robots do articulate? Are we more similar to robots than to frogs for example?

In this article we interpreted Polanyi's philosophy and came to the conclusion that certain machines do have tacit knowledge. But we also emphasized that this does *not* mean they are identical with living things.

This avoids a common problem in the philosophy of AI: that with the achievement of machine knowledge in an area previously dominated by humans, many think that human knowledge is successfully reproduced. This would indicate an ever-growing danger for the satisfactory demarcation of the human race from everything else—a development that is worrisome for many.

In our view the demarcation is clearer than ever. Thanks to Polanyi's philosophy, discussions of machine knowledge may continue.

6. Acknowledgement

This research was supported by the grant TÁMOP - 4.2.2.B-10/1--2010-0009 (and OTKA PD 83589).

Department of Philosophy and History of Science
Budapest University of Technology and Economics
mihaly.heder@filozofia.bme.hu,
daniel.paksi@filozofia.bme.hu

Notes:

1. p. 141-142: 'If I know how to ride a bicycle or how to swim, this does not mean that I can tell how I manage to keep my balance on a bicycle or keep afloat when swimming. I may not have the slightest idea of how I do this or even an entirely wrong or grossly imperfect idea of it, and yet go on cycling or swimming merrily. Nor can it be said that I know how to bicycle or swim and yet do *not* know how to co-ordinate the complex pattern of muscular acts by which I do my cycling or swimming. I both know how to carry out these performances as a whole and also know how to carry out the elementary acts which constitute them, though I cannot tell what these acts are. This is due to the fact that I am only subsidiarily aware of these things, and our subsidiary awareness of a thing may not suffice to make it identifiable.' p. 144: 'Such knowledge is ineffectual, unless known tacitly.'

2. See also: Paksi, 'Emergence and Reduction in the Philosophy of Michael Polányi Part I & II'. Part I. in *Appraisal*. Vol. 8. No. 2. 34-41. 2010. Part II. in *Appraisal*. Vol. 8. No. 4. 28-42. 2011
3. Of course, examining humans is not the only way of creating machines. It is possible to skip the first four steps in this list with artificial evolution for example.

References

Polanyi, Michael.

1953. Letter to Karl Polanyi (December 3, 1953), in English. Chicago: Box 17, Folder 2., as cited by Endre J. Nagy 'After *Polanyiana* Volume 5, Number 1, 1996, p. 77-100

1968b. 'Logic and Psychology'. *American Psychologist* 23. p. 27-43.

2010. 'Notes on Mind and Machine' (in Polanyi Archives). *Polanyiana* Volume 19, Number 1-2, p 97.

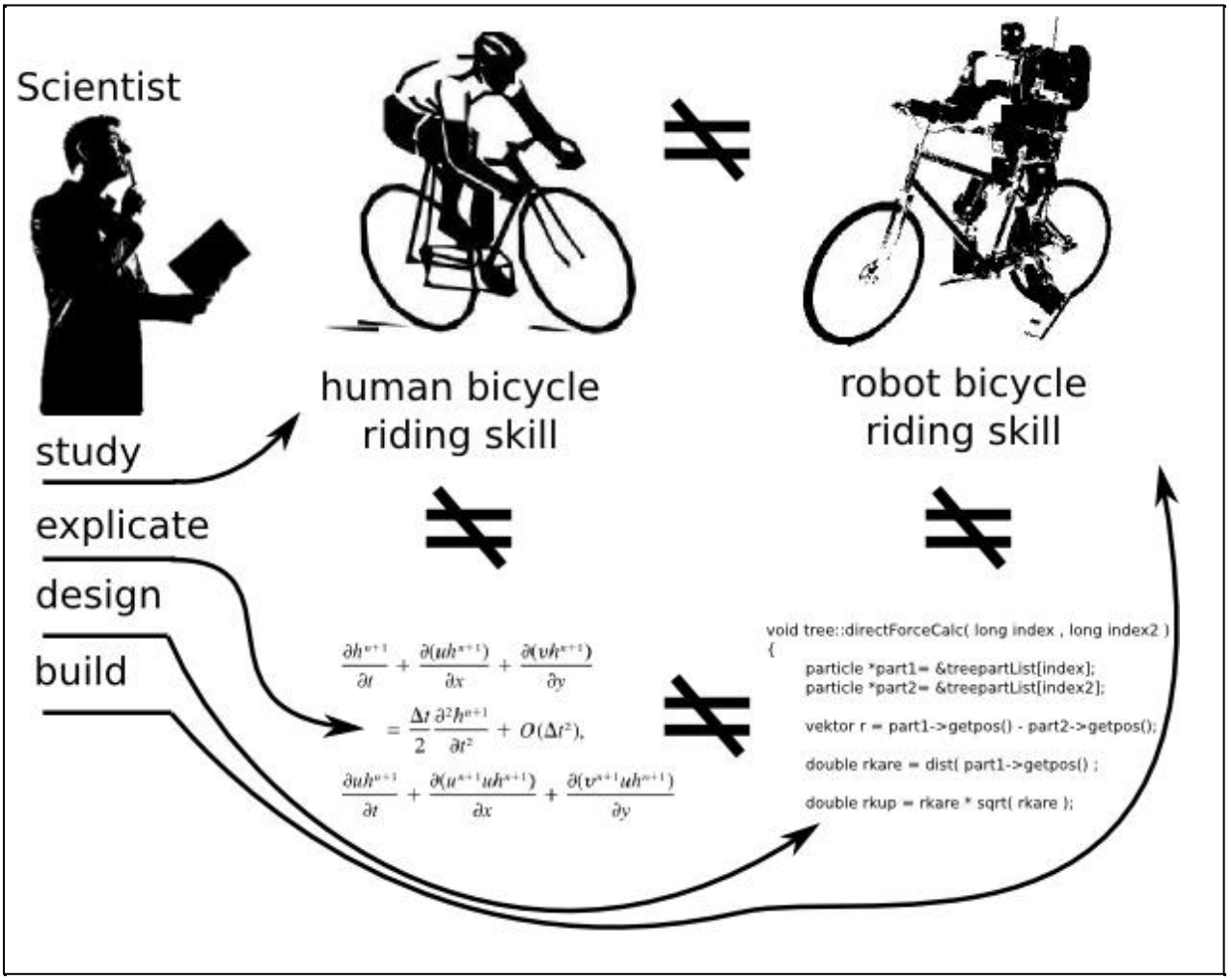

Fig. 1