The "Embodied Mind" Paradigm in Artificial Intelligence

Piotr Urbańczyk

About 10 years ago, Paul Dourish noticed that "computer science is based entirely on philosophy of the pre-1930's"[1]. He proposed rethinking this state of affairs and to enrich computer science with some more modern approaches to human cognition. In this paper I would like to show that his proposal is currently being implemented, especially in the domain of artificial intelligence (AI). My plan is to indicate that the same paradigm shift that was made in cognitive science is being carried out within AI.

1. Paradigm shift in cognitive science

Cognitive science is the interdisciplinary study of the mind and its operations. It includes research on intelligence and behaviour and comprises such disciplines as philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology. Its relationship with computer science, especially artificial intelligence (AI), is particularly varied and important. They were developed more or less simultaneously and one can easily indicate the mutual influence they have on one another.

Cognitive scientists apply various research methods and many of them are typical for 'ordinary' neuroscience, e.g., single-cell recording, neuroimaging (PET, fMRI) and lesion-studies. These methods are extended with various types of behav-

[1] P. Dourish, *Where the Action Is: The Foundations of Embodied Interaction*, MIT Press, Cambridge 2001, p. vii.

ioural experiments. Moreover, cognitive scientists often refer to comparative research concerning human brains and those of other primates. The unique aspect of cognitive science is based on the fact that those methods are used not only for examining neurons or neuronal structures, but also for examining the cognitive mechanisms of human beings. There is no doubt that the level of mental phenomena causes the most problems and it can be understood in various ways.

> Cognitive neuroscience sees psychological levels (conceptualized as, e.g. "cognition," "information processing," "representation," "computation") as the higher levels of description, to be explained by referring to the neural and neurocomputational mechanisms residing at the lower levels. In this view, psychological phenomena are not explanatorily autonomous, but neither are they eliminable – just like cytology is neither eliminable nor autonomous in the relation to biochemistry and molecular biology. Psychological properties are regarded as residing at a level of organization higher than neural properties, but nevertheless as being micro-based properties essentially in the same sense as other special-science properties.[2]

For that reason, the whole gamut of methods is supplemented with several assumptions which have arisen from the adoption of what can be called interpretative paradigms. They can be considered as an attitude or meta-theory that provide rules governing the construction of experiments, methods of interpretation of experimental data, basic objectives of research, methods of generation of scientific explanations, criteria of justification, understanding of basic concepts, e.g. the mind, and – last but not least – some anthropological and philosophical

[2] A. Revonsuo, *On the Nature of Explanation in Neurosciences*, [in:] *Theory and Method in the Neurosciences*, eds. P. Machamer, R. Grush, P. McLaughlin, University of Pittsburgh Press, Pittsburgh 2001, p. 56.

assumptions.[3] Such paradigms are computationalism, evolutionary psychology as well as the 'embodied-embedded mind' paradigm, to name the most popular.

The supporters of the computational approach used to interpret experimental data in terms of information processing. The level of mental phenomena is treated by them as strictly algorithmic (as a software), but implemented in the biological hardware.[4] The main objective of the computationalists is, above all, explanation through the discovery of computational mechanisms and the creation of cognitive architectures.

Evolutionary psychologists adopt the basic postulate of the computationalists concerning the psychological level, namely, the computability of the mind. This demand usually takes the form of the strong modular theory of mind, known as Massive Mental Modularity.[5] Evolutionary psychologists also emphasize the evolutionary origins of mind.[6] In their pattern of scientific explanation they usually refer to the adaptational advantages (increasing fitness) related to particular mental modules (mechanisms). In short, on the basis of evolutionary psychology, 'to explain something' means to show its adaptive function. For example, a typical evolutionary psychologist would say that the cognitive mechanisms of face recognition have arisen as an adaption which enabled the recognition of relatives (which is

[3] Cf. B. Brożek, *Philosophy in Neuroscience*, [in:] *Philosophy in Science. Methods and Applications*, eds. B. Brożek, J. Mączka, W.P. Grygiel, Copernicus Center Press, Kraków 2011, pp. 181–183.

[4] J.R. Anderson, *Methodologies for Studying Human Knowledge*, "Behavioral and Brain Sciences" 1987, no. 10, pp. 467–505.

[5] The forerunner of this approach is Jerry Fodor; cf. *idem*, *The Modularity of Mind*, The MIT Press, MA-London 1983.

[6] See *The Adapted Mind. Evolutionary Psychology and the Generation of Culture*, eds. J.H. Barkow, L. Cosmides, J. Tooby, Oxford University Press, NY-Oxford 1992.

crucial for kin selection) as well as recognition of the recipients of acts of altruism, from whom we expect reciprocation.

In turn, the representatives of the embodied-embedded mind theory definitely reject the postulates of the computability and modularity of mind. Although they treat the theory of evolution seriously, they consider the above scheme to be naive (i.e. a scheme in which to explain means to show the evolutionary adaptive function). Not all of the products of evolution have an adaptive nature (most of the products of evolution are by-products).

The embodied-embedded mind paradigm is largely based on the achievements of such sciences as applied linguistics (e.g., the theory of conceptual metaphors by George Lakoff), anthropology (e.g., the theory of the cultural origins of human cognition by Michael Tomasello), and also on the new achievements of neurobiology (e.g., the theory of mirror neurons and embodied simulation).[7]Generally speaking, this paradigm shows the enormous role played by the physical interactions between individuals and their social and cultural environment in the shaping of their cognitive abilities and mental states. Although the 'embodied-embedded mind' paradigm seems to be very fertile, it is also not free of assumptions which are difficult to test empirically.

The idea of embodiment comes from the phenomenologist Maurice Merleau-Ponty. He claims that representation is not created by the mind itself, but it is very closely connected to bodily perception and action. Moreover, it is always constructed

[7] Cf. G. Lakoff, *Women, Fire and Dangerous Things. What Categories Reveal about the Mind*, The University of Chicago Press, Chicago 1987; M. Tomasello, *The Cultural Origins of Human Cognition*, Harvard University Press, Cambridge 1999; V. Gallese, *Embodied Simulation*, "Phenomenology and Cognitive Sciences" 2005, no. 4, pp. 23–48.

by an embodied agent during one's interaction with the world.[8]A more up to date illustration of how our higher-order cognition can be traced back to its bodily basis was given by the above mentioned pair of George Lakoff and Mark Johnson. In their book *Philosophy in the Flesh*[9] they show that complex concepts in our minds can be mapped onto the domain of bodily orientation and our movement in space. More recently, some results in neurobiology indicate that the cerebral representation of what is happening to our body is important not only for the motorics of activities undertaken in the physical environment, but this mechanism is also involved in the formation of the more complex content of our minds. The process of mapping the body is intricate not only in the purely biological level – it also comprises all interactions between the body and the environment and, therefore, human activity in the physical and social setting.[10]

2. Good old-fashioned artificial intelligence

What does it mean that a computer or a robot is intelligent? How can we tell that it performs intelligent behaviour? What does it take to get a computer to engage in such activities? What is artificial intelligence (AI)? The answers to these questions depend on what we consider intelligent and what intelligence is in general. According to the very famous but general

[8] H.L. Dreyfus, *Intelligence without representation: Merleau-Ponty's critique of mental representation*, "Phenomenology and the Cognitive Sciences", 2002, no. 1 (4), pp. 367–383.

[9] G. Lakoff, M. Johnson, *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*, Basic Books, New York 1999.

[10] S. Gallagher, *Phenomenological and experimental contributions to understanding embodied experience*, [in:] *Body, Language and Mind. Vol. 1*, eds. T. Ziemke, J. Zlatev, R. Frank, R. Dirven, Mouton de Gruyter, Berlin 2007, pp. 241–263.

division made by Stuart Russell and Peter Norvig, there are four approaches to defining AI.[11] We can talk about:

	human likeness	rationality
thinking	Systems that think like humans	Systems that think rationally
acting	Systems that act like humans	Systems that act rationally

Several conclusions can be drawn from this table. Firstly, this division does not resolve the philosophical questions related to the topic. We still do not know how people think – there is an on-going struggle among cognitive scientists concerning human thinking. There is also some doubt concerning the right-hand side of the table – are systems that think or act rationally better than systems that think/act like humans. Human thought is often imperfect. If we defined rationality as Russell and Norvig did - i.e. as thinking according to the "laws of thought" like logic – computers would have a distinct advantage over imperfect human beings. We often form false beliefs, cannot derive a simple conclusion from a small set of premises and so on. Nevertheless, we can define rationality as goal-oriented behaviour and thinking. In this case, we could admit that even the behaviour of an "unintelligent" animal (e.g, an ant searching for food) is rational.

Hector J. Levesque[12] considers some of the examples of intelligent behaviour that computers can do, such as understanding natural language sentences, recognizing objects in

[11] S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, New Jersey 2003, p. 2.

[12] H. Levesque, *Thinking as Computation. A First Course*, MIT Press, Cambridge-London 2012.

a visual scene, planning courses of actions, solving recreational puzzles or playing strategic games. He concludes that those activities have something in common – when they are performed by people, they appear to require thought.

Does this allow us to claim that computers are "electronic brains"? Historically speaking, the human brain has often been modelled on the most advanced technology of the time such as clockwork, the steam engine or a telephone switchboard. Nowadays we find these models to be simplistic and misleading and they tell us very little about what our brains are, and how we do what we do with them. On the other hand, we used to compare our brains to computers. Why should we think that those machines are any different? How can we be sure that we will not end up laughing at such models?

What is thinking then? Consider the following sentence "The hat would not fit into the suitcase because it was too small" How do we figure out what "it" is in this sentence, that which is too small? Observe that there is nothing in the sentence itself that gives away the answer. We can replace "small" by "big" – "The hat would not fit into the suitcase because it was too big". What does "it" refer to now? We can determine what "it" refers to by utilizing what we already know about the sizes of objects and fitting one thing into the other. This is thinking and the example shows us what thinking looks like in action. Some people say that thinking is a biological process that happens in the brain, like digestion in the stomach or mitosis in cells. It is clearly biological process, because we are biological creatures. Nevertheless, some computer scientists, especially AI researchers (Levesque among them) would like to claim that thinking can be usefully understood as a *computational* process.

In turn, a computation is certain form of the manipulation of symbols. We take strings of symbols, break them apart, compare them, and reassemble them according to a recipe called a *procedure*. Levesque observes that computers do not have to understand what the symbols stand for or why the manipulations are correct. The symbols can be manipulated completely mechanically and still end up producing significant and interesting results. We can get computers to perform a wide variety of very impressive activities precisely because we are able to describe those activities as a type of symbol manipulation that can be carried out purely mechanically.[13]

This paradigm is called good old-fashioned artificial intelligence (GOFAI). Over the last 40 years some computer scientists have tried to show some of the limitations of this paradigm in AI. Perhaps the most influential papers were written by Hubert Dreyfus[14] and Rodney Brooks[15]. A very good overview of this discussion was given by Michael L. Anderson[16].

3. Embodied cognition in artificial intelligence

Traditional AI consists of understanding intelligence in terms of thought and reason and involving representations and high-level cognitive skills like planning or problem-solving. According to Brooks this approach is too shallow.[17] While preferring studying intelligence "from the bottom up" he emphasizes the evolutionary origin of this human capacity. Intelligence was developed to meet our needs and to deal with an environ-

[13] *Ibidem*, p. 10.

[14] H. L. Dreyfus, *What Computers Can't Do: A Critique of Artificial Intelligence*, Harper and Row, New York 1972.

[15] R. Brooks, *Cambrian Intelligence: The Early History of the New AI*, MIT Press, Cambridge 1999.

[16] M. L. Anderson, *Embodied Cognition: A field guide*, "Artificial Intelligence", 2003, no. 149, pp. 91–130.

[17] See R. Brooks, *op.cit.*, p. 134.

ment. He also recalls the continuity between humans and other animals.

One of the difficulties this approach has to face is a problem with optimization. Systems built within this approach are insufficiently dynamic. This is because the framework they use, called by Brooks SMPA (sense-model-plan-act framework). Mostly such robots operate in an environment specially engineered for them. They sense the world and build a model of it. Then they can ignore the actual environment and produce a plan of action based merely on a model. But the actual world is quite dynamic. When the actual world changes, the system has two options – either run the SMPA procedure again or execute an inappropriate action. What it would finally do depends on its sensitivity. Although nowadays microprocessors are very fast and can construct SMPA systems that operate in a real-world environment in a realistic time-scale, the SMPA framework is too expensive for Brooks by its nature and therefore biologically implausible.[18] According to him it would be better to use the world as its own model.

Another essential difficulty of the SMPA systems is the relevance problem as pointed out by Anderson.[19] A system or a robot does not have to update its plans and actions every time the world changes. It has to do that only in the face of *relevant* change, i.e. one that could prevent it achieving its goal. For this reason, it has to be equipped with a program that is able to indicate what counts as relevant and what does not. Anderson illustrates this practical problem with a funny quotation from Dennett:

> Back to the drawing board. 'We must teach it the difference between relevant implications and irrelevant implications', said

[18] See R. Brooks, *op.cit.*, pp. 136–137. For the discussion see M. L. Anderson, *op.cit.*, pp. 95–97.

[19] See M. L. Anderson *op.cit.*, pp. 97–98.

> the designers, 'and teach it to ignore the irrelevant ones'. So they developed a method of tagging implications as either relevant or irrelevant to the project at hand, and installed the method in their next model, the robot-relevant-deducer, or R2D1 for short. When they subjected R2D1 to the test that had so unequivocally selected its ancestors for extinction, they were surprised to see it sitting, Hamlet-like, outside the room containing the ticking bomb, the native hue of its resolution sicklied o'er with the pale cast of thought, as Shakespeare (and more recently Fodor) has aptly put it. 'Do something!' they yelled at it. 'I am', it retorted. 'I'm busily ignoring some thousands of implications I have determined to be irrelevant. Just as soon as I find an irrelevant implication, I put it on the list of those I must ignore, and.. .' the bomb went off.[20]

Both Anderson and Brooks make one comment here– the root of the relevance problem is a problem of representations. There is no context-free, absolute world model. Every representation is selective – it is oriented towards a specific goal. The problem of relevance requires representations to be more specific and limited.

4. What does it mean for AI?

According to Anderson, the failure of SMPA framework leads to "more reactive, agent-relative model of real-world action". He writes:

> The above problems seem to suggest their own solution: shorter plans, more frequent attention to the environment, and selective representation. But the logical end of shortening plan length is the plan-less, immediate action; likewise the limit of more frequent

[20] D. C. Dennett, *Cognitive wheels: The frame problem of AI*, [in:] ed. C. Hookaway, *Minds, Machines and Evolution*, Cambridge University Press, Cambridge 1984, pp. 129.

> attention to the environment is constant attention, which is just to use the world as its own model. Finally, extending the notion of selective representation leads to closing the gap between perception and action, perhaps even casting perception largely in terms of action.[21]

The author himself notices that this idea suggest a notion known from embodied cognition (or better – cognitive science done in embodied mind paradigm), namely that our perpetual field is always at the same time our action field. All objects in our close environment are perceived in relation to position and orientation to one's own body and then known in terms related to her current possibilities and future actions. On the basis of the embodied mind paradigm, representations (especially visual ones) are strongly associated with motor imagery. Grasping a cup or catching a moving ball could serve as examples. Movement and location of those objects are always represented according to the body and then used to simulate the movement of the hand. Furthermore, Anderson's idea is similar to the theory of direct perception understood in terms of affordances – we perceive things with their functionalities, availability to certain interventions, their calling for actions.[22]

It could be stated that this considerations have a rather philosophical (or theoretical) character and has nothing to do with applied AI. Nevertheless, the approach proposed by Anderson, Brooks and others initiated a research program that has had an

[21] M. L. Anderson *op.cit.*, pp. 99.

[22] See R. Grush, *Skill and Spatial Content*, "Electronic Journal of Analytic Philosophy", 1998, no. 6, online access (20.02.2014): http://mind.ucsd.edu/misc/ejap/ejap_6_6_Grush.html; J. J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, New York 1979; W. J. Clancey, Situated Cognition: *On Human Knowledge and Computer Representations*, Cambridge University Press, Cambridge 1997.

impact on robot design. A great number of such robots[23] (together with a good theoretical account) was described by Andy Clark[24].

Many embodied approach supporters point out another two interesting aspects. I will only draft them here. Firstly, the embodied approach in AI pays a great deal of attention to "new ways of interacting with computers, ways that are better tuned to our needs and abilities"[25]. Generally speaking, it has contributed to user-oriented interface design. Secondly, it has initiated a new way of thinking about how we can employ the largely unused power of computers to extend our natural human skills and abilities.[26]

5. Conclusion

There is no doubt that understanding thinking as a purely computational process is highly controversial. It should be pointed out, however, that similar conjecture was the basis for today's cognitive science. The forerunners of cognitive science assumed that the mind has a computational character and serves as *software* implemented on biological *hardware*, i.e. the brain. We can say that it was the first paradigm of cognitive science. It provided rules for the construction of experiments, methods of interpretation of experimental data, the basic objectives of research, methods of generation of scientific explanations, methods of understanding basic concepts of for example

[23] The robot *Cog* of Rodney Brooks made within the project conducted at MIT as well as CB2 – biomimetic robot made at the University of Osaka Minoru Asada were probably the most spectacular examples of these. Nowadays there are many humanoid robots made around the world by various teams and for various purposes.

[24] See A. Clark, *Supersizing the Mind. Embodiment, Action, and Cognitive Extension*, Oxford University Press, New York 2008, especially chapter 1.

[25] P. Dourish, *op. cit.*, p. 2.

[26] Cf. A. Clark, *op. cit.*

the mind and – last but not least – some anthropological and philosophical assumptions. We cannot deny that this paradigm seemed to be very fruitful. Evolutionary psychologists adopted this assumption (namely, the computational character of the mind), but they emphasized its evolutionary origins. This demand usually takes the form of the strong modular theory of mind, known as Massive Mental Modularity. Nowadays theories constructed within the so-called "embodiment mind" paradigm are becoming more popular. The supporters of such theories definitely refuse the postulates of the computability (and modularity) of the mind. This way of thinking appeared also in the domain of computer science. The enthusiasts of embodied cognition (EC) in artificial intelligence have tried to indicate some of the problems that GOFAI copes with.

The idea underlying GOFAI consists of taking the process of thinking separately from its brain and bodily basis. It is unacceptable in the EC approach. According to its supporters, our intelligence was developed in the process of evolution as a result of our bodily interactions with the environment. In the light of this struggle, the study of flight would be useful analogy. About a century ago, people were faced with the task of building a flying machine and they tried to do so according to two different strategies. One of them tried to understand how animals like birds could fly, carefully studying their feathers and muscles etc. and constructing machines to emulate them. The others simply wanted to build machines that were capable of flight. They studied aerodynamics – the principles of flight applicable to anything. While the EC approach tries to artificially reconstruct human-like intelligence, GOFAI tries to apply rational "rules of thought". Although, in my opinion, thinking is not merely a computational process, I am sure that there is nothing wrong with adopting a weaker conjecture. Namely, it (sometimes) can be considered as computation and some results of thinking can be achieved with computational, or even mechanical methods. But it does not cover all aspects of intelligence. In the case of flying machines, the second strategy turned out to be more successful. In the case of thinking machines, it is still very fruitful, but I believe that the first one would be even better.