

A Short Note on Emergence of Computational Robot Consciousness¹

Jozef Kelemen

Institute of Computer Science, Silesian University, Opava, Czech Republic,
and VSM College of Management, Bratislava, Slovak Republic
kelemen@fpf.slu.cz

Abstract: A way is sketched how to answer the question about the computational power supposed behind the consciousness, esp. the computational robot consciousness. It is illustrated that a formal model of the possible functional architecture of certain type of robot enables to satisfy the test of emergence proposed earlier.

Keywords: Robots, Machine Consciousness, Emergence, Eco-Grammar Systems

1 Introduction

This article in certain extent enlarges the existing spectrum of consciousness-oriented computer science opinions proposing a possible way of how to treat consciousness from the positions of the theory of computation, especially from the point of view of the theory of formal grammars and languages, more specifically, of the theory of so called *eco-grammar systems* as presented in (Csuhaj-Varju et al., 1997).

Consciousness will not need to be programmed in. They will emerge, stated R. Brooks (1999, p. 185), one among the leading specialists of the present days artificial intelligence and advanced robotics research. Our main goal consists in arguing – from a theoretical computer science point of view – for the *possibility that* consciousness will emerge. First, we will try to show some *views of consciousness* which seem to be relevant for our computationalistic treatment of the topic. Then we will focus to another relevant matter – to the *nature of emergent phenomena*, and to the *phenomenon of emergence*.

¹ The author's research is supported by the Czech Ministry of Education grant MSM 4781305903, and by the Gratex International, Inc., Bratislava, Slovakia. The present article is based on the author's previous contribution (Kelemen, 2006a) delivered at the 15th International Workshop on Robotics in Alpe-Adria-Danube Region, RAAD 2006 (Balatonfüred, Hungary, June 15-17, 2006).

In order to connect the contents of the centents of the just mentioned sections, we will continue with sketching two fundamental *approaches to the architecture of robots*. The sketched architectures will be then analysed from the perspective of their *computational power*. We will recognize at least one fundamental difference between the two basic architectural approaches to robot construction, and we will close the article with expressing the discovered difference in the conceptual framework of advanced robotics.

2 Consciousness

A. Zeman – approaching the study of consciousness from positions of a neurophysiologist with a considerable strong philosophical background – writes on three basic meanings of consciousness, all related with knowledge: *Being awake, our first sense of cons-ciousness*, he writes, *is a pre-condition for acquiring knowledge of all kinds. Once awake, we usually come – he continues – by knowledge through ex-perience, the second sense of cons-ciousness. The knowledge we gain is then ‘conscious’ in the third sense we distinguished*, he completes his analysis (Zeman, 2002, p. 36).

From another position – from the position typical for the fields of artificial intelligence and cognitive science – is the subject of consciousness treated e.g. by P. O. Haikonen. As a crutial for forming the consciousness he recognize the phenomenon of perception of the self. *What we actually see is only the projected image on the retina, what we actually hear is the vibration of the eardrums, so why don’t we just perceive these as such, percepts originating at the senses or originating at the related sensory nerve endings? How can we possibly perceive the situation being anything else?* he asks (Haikonen, 2003, p. 71).

The above questions formulated by Haikonen lead to the concept of some kind of “internal” (mental) representation of the “external” (physical) stimuli sensed by machines, and in consequences to the familiar mind/body problem of the philosophy of mind. Haikonen’s position is explained by an example (Haikonen, 2003, pp. 248-249) as follows: *The operation of the signals in the cognitive machine can be compared to radio transmission where a carrier signal is modulated to carry the actual radio signal. The carrier wave is what is received, yet what is detected is the modulation, the actual sound signal that is in causal connection to the original physical sound via a microphone. We do not hear the carried signal even through without it there would be no music. Thus it is possible to perceive carried informa-tion without the perception of the material basis of the carrier.*

In (Holland, Goodman, 2003), an analysis is given concerning the role internal symbolic representations of the robots environment and their own capabilities of

robots – as the term representation is used in traditional artificial intelligence research – to robots abilities to act in the their outer environments, and the abilities of robots to construct and use their internal symbolic representations are connected with the phenomenon of the robots consciousness.

3 Emergence

The traditional and most wide informal definition of emergence has been formulated e.g. in (Holland, 1998, pp. 121-122): Emergence is, according him „... *a product of coupled, context-dependent interactions. Technically these interactions, and the resulting system, are nonlinear: The behavior of the overall system cannot be obtained by summing the behaviors of its constituent parts... However, we can reduce the behavior of the whole to the lawful behavior of its parts, if we take nonlinear interactions into account*“.

In (Searle, 1992) at least two interpretations of the concept of emergence are distinguished: The first one Searle calls *emergence1*. This kind of emergence refers that a higher order feature of a system can be understood by a complete explication of the parts of a system and their interpretation. The more adventurous conception of emergence Searle calls *emergence2*. A feature of a system emerges in this way if it has causal powers that cannot be explained by the parts of a system and their interactions.

The consequence of the conception of *emergence2* for the consciousness is the following: If consciousness is of the *emergence2* type, that it could cause things that could not be explained by the causal properties of the neuronal networks. A serious problem arising from the *emergence2* – called in (Van Gulick, 2001) as *radical kind of emergence* – consists in making the physicalist view of consciousness problematic: *If [...] systems could have causal powers that were radically emergent from the powers of their parts in the sense that those system-level powers were not deter-mined by the laws governing the powers of their parts, then that would seem to imply the existence of powers that could override or violate the laws governing the powers of the parts*, states (Van Gulick, 2001, p. 18).

The emergent nature – we hope the radical one – of phenomena appearing in complex systems we may test using the so-called *test of emergence* proposed in (Ronald et al., 1999). The requirements putted onto systems in which the emergence of some phenomenon appears are the following (Roland et al., 1999):

Design. The designer designs the systems by describing *local* in-teractions between components in a language L_1 .

Observations. The observer describes global behaviors of the running system using a language L_2 .

Surprise. The language of design L_1 and the language of observation L_2 are distinct, and the causal link between the elementary interactions program-med in L_1 and the observations observed in L_2 is non-obvious.

4 Robots

Intuitively we feel that any robot consciousness necessarily requires attentiveness and emotionality. This opinion is expressed clearly in first dreams on robots in artistic works; cf. e.g. (Horakova, Kelemen, 2008), but also in numerous theoretical studies rooted in the computational and engineering approaches, e.g. in an attempt to a formal axiomatic definition of consciousness, an approach by which we will inspired in this section (Aleksander, Dunmall, 2003). According their opinion, being a *conscious agent* means – intuitively and roughly speaking – to have some kind of agent’s *private sense* of an outer world, of a self in this world, of self’s contemplative planning of when, what and why to act, of self’s own inner emotional states Moreover it means also the conscious agents ability to include the self’s *private sense* into all of its above mentioned functional capabilities. But how incorporate it into agents?

All the real experimental robots which work with internal symbolic representations of their outer environment have from our perspective one important common feature: At least in certain extent, their behavioral, representational, and decision making capacities are based on the abilities of the present day computers to execute more or less complicated computations. In a theoretical level, these computations might be reduced into the form of the theoretical abstraction of computational processes known as *Turing-computations*, so computations performed by the abstract universal Turing machine. The *Turing machine* – with respect of their computational power in the machines oversimplified environments of symbols on a tape, and a head going one step left or right and rewriting the symbols according simple instructions sequences – differs in an important sense very significantly from the real embodied robots situated in dynamically changing physical environments, and interacting with these environments very massively in many different ways.

However, there exists a largely accepted hypothesis in theoretical computer science – the *Church-Turing hypothesis*; see e.g. (Cleland, 1993) for some discussion concerning it – according which, very roughly speaking, all what is intuitively in certain sense computable (so, transformable from certain inputs into certain outputs according precisely defined and exactly executed sequences of rules – according computer programs) is computable by the Turing machine.

Especially interactions are very appealing for re-consideration of the form of a “computation” performed by agents or robots, and for drawing perhaps new boundaries between what we consider as computable and what as non-computable.

In present day theoretical computer science there are numerous efforts to demonstrate that the notion of computation might be enlarged beyond the traditional boundaries of the Turing-computability. In (Burgin, Klinger, 2004) it is proposed to call algorithms and automata that are more powerful than Turing machines as *super-recursive*, and computations that cannot be realized or simulated by Turing machines as *hyper-computations*.

Turn now our attention toward the provoking notion of the *private sense* related to robots. To have the *private sense* means – metaphorically speaking, for more details see (Kelemen, 2006b) – to have an ability of a given robot to consider itself as another robot identical with it, and to consider this type of “schizophrenia” in the work of other functions which characterize our real robot. This type of recursion is might be extremely complicated for expressing it in the frame of the traditional paradigm of one-processor computation. It requires at least some suitable framework for dealing with behaviors that appear thanks to interrelations between individually autonomous robots.

The appearing situation insinuates the framework of considering a conscious robot as a system consisting in more than one agent, so in a form of a multi-agent systems. The robot’s private sense is perhaps an emerging product of interactions of several other robots functional modules. Perhaps the conscious behavior of such a robot might be then described as a phenomenon, which emerges – in the above cited sense proposed in (Holland, 1998) – from interactions of traditionally computable behaviors of simpler constituting parts of it, and has the form of a hyper-computation. Let us try to demonstrate in the next section how it is possible to proceed in this way.

5 Emergent Computation

There are several different approaches to study the emergent computational power of interacting systems. In this Section we will sketch a formal model of robots with functional components producing a rule-governed Turing-computable behaviors each, but producing – as a whole – a behavior which does not be generated traditionally by any Turing-equivalent generative device, so which requires the generative power of hyper-computation. We will consider in this role the so-called eco-grammar systems. First, we introduce in a few words this model, presented originally in (Csuhaj-Varjú et al., 1997).

According (Csuahaj-Varju et al., 1997), an *eco-grammar system* Σ consists, roughly speaking, of

- a finite alphabet V ,
- a fixed number (say n) of components evolving according sets of rules P_1, P_2, \dots, P_n applied in a parallel way as it is usual in L-systems (Rozenberg, Salomaa, 1980), and of
- an environment of the form of a finite string over V (the states of the environment are described by strings of symbols w_E , the initial one by w_0).
- the functions φ and ψ which define the influence of the environment and the influence of other components, respectively, to the components (these functions will be supposed in the following as playing no roles, and will not be considered in the model of eco-grammar systems as treated in this article).

The rules of components depend, in general, on the state (on the just existing form of the string) of the environment. The particular components act in the commonly shared environment by sets of sequential rewriting rules R_1, R_2, \dots, R_n . The environment itself evolves according a set P_E of rewriting rules applied in parallel as in L systems.²

The evolution rules of the environment are independent on components' states and of the state of the environment itself. The components' actions have priority over the evolution rules of the environment. In a given time unit, exactly those symbols of the environment that are not affected by the action of any agent are rewritten.

In the EG-systems we assume the existence of the so-called *universal clock* that marks time units, the same for all components and for the environment, and according to which the evolution of the components and of the environment is considered.

In (Wätjen, 2003) a variant of EG-systems without internal states of components is proposed and studied. The fixed number of components of the so-called teams of components in EG systems originally proposed in (Csuahaj-Varju, Kelemenova, 1998) is replaced by a dynamically changing number of components in teams. As the mechanism of reconfiguration, a function, say f , is defined on the set N of integers with values in the set $\{0, 1, 2, \dots, n\}$ (where n is the number of components in the corresponding EG-system) in order to define the number of components in teams. For the i -th step of the work of the given EG-system, the function f relates a number $f(i) \in \{0, 1, 2, \dots, n\}$. The subset of the set of all components of thus EG-system of the cardinality $f(i)$ is then selected for executing the next derivation step of the EG system working with Wätjen-type teams. So, Wätjen, roughly speaking, proved that there exist EG-systems such that if f is (in the traditional sense) non-recursive function, then the corresponding EG-system generates a non-recursive (in fact a super-recursive) language.

² So, the triplet (V, P_E, w_E) is (and works as) a Lindenmayer-system.

The *emergent nature* of the behavior (language) generated by the above described EG system is – applying the above mentioned test of emergence – rather clear: The components of a given EG system generate *recursive languages* each. Recursive languages play, in the context of the above cited *emergence test* the role of the *designer language*. The local interactions of the components of the systems are given only. But *surprisingly*, the whole system generates a *non-recursive language* (behavior), which is, as the *language of the observer* of the system behavior, substantially, surprisingly, different, from computational positions, from the *language of the designer*.

Conclusions

We saw that there exist formalized systems set up from decentralized components with higher computational power as Turing machines have. There are no principal reasons to reject the hypothesis that it is possible to construct real robots as certain kind of implementations of these formalized systems. If we include into the functioning of such robots the activation of their functional modules according a non-recursive (in Turing sense) computation, the behavior of the agents might be non-recursive. We suppose that this situation may appear inif some of the functional partsof the robots are swich on or off on the base of the random behavior of the robots environments, for instance. So we exclude the situation when a computer simulation of randomness are included into the functional architecture of robots. Rather, we suppose the randomness appearing in the environment, a randomness which follows from the ontology of robots situated in their environments. More about this can be found in (Kelemen, 2005a) and (Kelemen, 2005b).

The ontological randomness might be caused by different reasons – by inprecise work of sensors and actuators of robots, by erroneous behavior of their hardwired or software parts, by non-determinism of the behavior of the environment, etc. All these influences may be reflectedin the specific behavior of the robots and we cannot reject the hypothesis that just these kind of irregularities cause also the phenomenon called robot consciousness.

So, going back to the Brooks' opinion from the beginning of this contribution: Consciousness will not need to be programmed in. They will emerge. The goal of this contribution was not to prove that *it will emerge* sometime during the course of time, but to prove that *it may emerge* form the functional-computational structure and from properties of robots, and their massive interactions with their complicated unpredictable behaving environments.

References

- [1] Aleksander, I., Dunmall, B.: Axioms and Tests for the Presence of Minimal Consciousness in Agents. In: *Machine Consciousness* (O. Holland, Ed.). Imprint Academic, Exeter, 2003, pp. 7-18

- [2] Brooks, R. A.: *Cambrian Intelligence*. The MIT Press, Cambridge, Mass., 1999
- [3] Burgin, M., Klinger, A.: Three Aspects of Super-Recursive Algorithms and Hyper-Computation or Finding Black Swans. *Theoretical Computer Science* 317 (2004) 1-11
- [4] Cleland, C. E.: Is the Church-Turing Thesis True? *Minds and Machines* 3 (1993) 283-312
- [5] Csuhaj-Varju, E., Kelemen, J., Kelemenova, A., Paun, Gh.: Eco-Grammar Systems – a Grammatical Framework for Lifelike In-Teractions. *Artificial Life* 3 (1997) 1-28
- [6] Csuhaj-Varju, E., Kelemenova, A.: Team Behaviour in Eco- Grammar Systems. *Theoretical Computer Science* 209 (1998) 213-224
- [7] Haikonen, P. O.: *The Cognitive Approach to Conscious Machines*. Imprint Academic, Exeter, 2003
- [8] Holland, J. Emergence. Holland, J. H.: *Emergence*. Addison-Wesley, Reading, Mass., 1998
- [9] Holland, O., Goodman, R.: Robots with Internal Models – a Route to Machine Consciousness? In: *Machine Consciousness* (O. Holland, Ed.) Imprint Academic, Exeter, 2003, pp. 77-109
- [10] Horakova, J., Kelemen, J.: The Robot Story – why Robots were Born and How They Grew Up. In: *The Mechanical Mind in History* (Ph. Husbands, O. Holland, M. Wheeler, eds.). The MIT Press, Cambridge, Mass., 2008, pp. 283-306
- [11] Kelemen, J.: May Embodiment Cause Hyper-Computation? In: *Advances in Artificial Life, Proc. ECAL 05* (M. S. Capcarrère et al., eds.). Springer-Verlag, Berlin, 2005a, pp. 31-36
- [12] Kelemen, J.: On the Computational Power of Herds. In: *Proc. IEEE 3rd International Conf. on Computational Cybernetics* (I. J. Rudas, Ed.). Budapest Polytechnic, Budapest, 2005b, pp. 269-273
- [13] Kelemen, J.: Computational Robot Consciousness – a Pipe Dream or a (Future) Reality? In: Proc. 15th International Workshop on Robotics in Alpe-Adria-Danube Region, RAAD 2006 (I. Rudas, Ed.). Budapest Tech, Budapest, 2006a, pp. 234-240
- [14] Kelemen, J.: Agents from Functional-Computational Perspective. *Acta Polytechnica Hungarica* 3 (2006b) pp. 37-54
- [15] Ronald, E. M. A., Sipper, M., Capcarrère, M. S.: Design, Observation, Surprise! A Test of Emergence. *Artificial Life* 5 (1999) 225-239
- [16] Rozenberg, G., Salomaa, A.: *The Mathematical Theory of L Systems*. Academic Press, New York, 1980

- [17] Searle, J.: *The Rediscovery of the Mind*. The MIT Press, Cambridge, Mass., 1992
- [18] Van Gulick, R.: Reduction, Emergence and Other Recent Options on the Mind/Body Problem. In: *The Emergence of Consciousness* (A. Freeman, Ed.) Imprint Academic, Thorverton, 2001, pp. 1-34
- [19] Wätjen, D.: Function-Dependent Teams in Eco-Grammar Systems. *Theoretical Computer Science* **306** (2003) 39-53
- [20] Zeman, A.: *Consciousness – A User’s Guide*. Yale University Press, New Haven, 2002