Co-evolutionary robotics and the problems of autonomy

Willem (Pim) F.G. Haselager
Nijmegen Institute for Cognition and Information, Nijmegen, The Netherlands
UNESP, Marilia, SP, Brazil
w.haselager@nici.kun.nl

Abstract

Much recent research in cognitive science has focused on the role of self-organization and emergence in the development of robots. Especially in the area of co-evolutionary robotics, the question is whether the role of the programmer in the developmental process can be diminished to such an extent that the resulting robots can be said to be autonomous. In this paper I will suggest that two problems of autonomy have to be distinguished. The first, technical, problem is related to driving the designer out of the robot. The second, hard, problem concerns the question when systems can be said to pursue goals that are truly their own. I argue that a deeper understanding of the role of the body in action, and its continuous coupling with the control system is essential in order to make progress in relation to the hard problem of autonomy.

1. Introduction

When one demonstrates a computer model or a robot to a general audience, one of the first questions that one will encounter is: ”but isn’t this model just doing what it is told to do?” Normally this question implies that the computational system is not genuinely engaging in the cognitive or behavioral acts that the demonstrator was talking about, but instead is obeying, in a puppet-like fashion, the programmer who is pulling (or pulled in advance) the strings. This objection is as old as AI [16], but with the increasing presence of robots in our daily environment, both at work and at home (e.g. Sony’s AIBO) the question regarding the autonomy of robots (not to mention the accompanying question regarding who is responsible for the consequences of their behavior) will become increasingly urgent.

In this paper I will investigate how recent work in (co-)evolutionary robotics is related to autonomy. On the basis of this investigation I will argue that two different versions of the problem of autonomy can be distinguished: a ‘technical’ and a ‘hard’ version. I will suggest that the ‘hard’ problem of autonomy requires a deepening of our understanding of the role of the body in the emergence of autonomous action.

2. Autonomy

Autonomous agents are able to operate under all reasonable conditions without recourse to an outside designer, operator or controller and to be capable of handling unpredictable events in an unstructured environment or niche, while pursuing their goals. The phrase ‘handling unpredictable events in an unstructured world’ needs to be included in order to rule out pre-configured systems operating blindly in a completely predetermined environment. Autonomy, in my view, does not imply that the agent needs to be conscious of its actions or the significance of its interaction with the environment. As Dennett points out: “At a certain point self-replicating macro-molecules cross the difference between just having effects and performing actions. Even though they do not know what they do, they do act. The quasi-agency involves purposive hustle and bustle, yet there is nobody home.” [4, p.21].

In relation to the aspect of being without recourse to an outside designer, operator or controller (i.e. independence), it is important to realize that autonomy does not constitute a dichotomy but rather a *continuum* between complete dependence and complete independence. Indeed, one may question whether cases of complete independence genuinely exist, since there is always some reliance of the system on the environment and/or other agents (e.g. [13]). Therefore, instead of asking if robots can be autonomous, I suggest to investigate the question to *what extent* robots can be autonomous.

The concept of having goals is complicated, and I will have more to say about it later. For the moment, then, autonomy will be taken here in the meaning of acting independently, either consciously or subconsciously, in an unpredictable and dynamic environment, in order to sustain one's existence (e.g. by maintaining sufficient energy, avoiding damage, etc.) and to achieve any further goals of one's own.

3. Robotics

In relation to autonomy, robotics constitutes a more promising approach in comparison with traditional AI (based on the idea of internal knowledge processing) because of its emphasis on the importance of the body and the environment for cognition. Robots are interesting because they are physically embodied and they have to actively interact with an uncertain and changing environment in order to pursue whatever goals they have (received). Robots, more than the traditional disembodied and non-embedded computer models, are candidates for instantiating a degree of agency and autonomy in virtue of their actual engagement with the world.

Another, more intuitive, reason for considering robots as prima facie good candidates for autonomy is that it is sometimes hard to avoid *sympathizing* with them. At least sometimes some robots seem to be struggling towards the attainment of some goal, trying to overcome obstacles in a quest for something or other. Some of their reactions to events seem to be based on their history of goal-directed interactions with the environment. Sympathy, of course, is no substitute for sound argumentation, but it does provide incentive to investigate the question whether the seemingly purposive behavior of robots is just 'mere seeming' or an indication of some degree of underlying autonomy.

Finally, much of the behavior of robots seems to be not predetermined but emergent, which, in the context of robots, can be understood as unprogrammed functionality [3, p.114]. The behavior of the robot is not directly controlled, or programmed for in a straightforward way, but arises out of the interactions between a limited number of components that can be substantially different in their properties and action possibilities. Clark gives the example of simple behavioral dispositions (tend towards the right, bounce back when touching something) in a robot that, under the right circumstances, could lead to emergent behavior such as wall following.

Still, the question remains: to what extent can robots be said to be autonomous? After all, they are created for the explicit purposes that their designers had in mind. Are their bodies, minds and behaviors not overly dependent on prior tinkering by developers and programmers? When a robot does something, does it do so at its own accord, or merely because it has been set up to do so?

4. Evolutionary robotics

In response to questions such as these, there have been continuous efforts to drive the programmer and developer 'out of the robot' as much as possible. Scientists working in the area of evolutionary robotics (e.g. [11]) have been at the frontier of this enterprise. Evolutionary robotics can be defined as: "the attempt to develop robots and their sensorimotor control systems through an automatic design process involving artificial evolution." [10, p.167].

Artificial evolution involves the use of genetic algorithms. The 'genotypes' of robots are represented as bits that can code their morphological features as well as the characteristics (such as weights and connections of a neural network) of their control systems. A fitness formula determines candidates for reproduction by measuring the success of the robots on a specific task. The genotypes of the selected robots are then subjected to crossover with other genotypes and further random mutation, giving rise to a new generation of robots. According to Nolfi [10, p.167-168], the organization of the evolving systems is the result of a self-organizing process, and their behavior emerges out of the interactions with their environment. Therefore, evolutionary robotics is relevant to the topic of autonomy since there is less need for the programmer and/or designer to 'pull the strings' to shackle the autonomy of the evolving creatures, because the development of robots is left to the dynamics of (artificial) evolution.

Artificial evolution is far from straightforward, however, and usually requires an extensive amount of preparation before the evolutionary process can take off. As Nolfi [10, p.179] points out: "In principle (…) the role of the designer may be limited to the specification of a selection criterion. However, (…) in real experiments the role of the designer is much greater than that: In most of the cases the genotype-to-phenotype mapping is designed by the experimenter; several parameters (e.g. the number of individuals in the population, the mutation and crossover rate, the length of the lifetime of each individual, etc.) are determined by the experimenter; and in some cases the architecture of the controller is also handcrafted. In theory, all these parameters may be subjected to the

evolutionary process; however, in practice they are not."

Moreover, it is not unusual to find that the designer not only pre-arranged the evolutionary process, but also interfered directly with its course in order to solve thorny issues such as local minima and the bootstrap problem. Local minima arise when after a certain amount of progress in relation to the performance of a task, new generations stop improving and the evolving robots get stuck in a sub-optimal performance. The bootstrap problem involves how to get beyond the starting point when the individual robots of the initial generation are unable to differentiate themselves in relation to the task because they all score zero, so that there is no way to select the 'best' or least worst individuals. In relation to problems such as these, Nolfi [9] reports, for instance, having to change the fitness formula by adding elements that cause reward for in itself not very meaningful behavior, and increasing the number of encounters with relevant stimuli. Often, then, "some additional intervention is needed to canalize the evolutionary process into the right direction." [9, p.196], keeping the designers well in control of their robots, even if the strings to which the robots are tied may be less visible.

A way of further reducing the role of the designer is to let the performance criteria take care of themselves, eliminating the need to pre-specify or modify them. This can be achieved through the co-evolution of different types of robots. As Nolfi [10, p.175] suggests: "A more desirable solution to the bootstrap problem (…) would be a self-organized process capable of producing incremental evolution that does not require any human supervision. This ideal situation arises spontaneously in co-evolving populations (i.e. in the case of competing populations with coupled fitness, such as predator and prey). Each co-evolving population may progressively produce more complex challenges for the other population. As a consequence (…) competing populations may reciprocally drive one another to increasing levels of complexity by producing an evolutionary 'arms race'".

For example, the weights of neural networks controlling the behavior of khepera robots, representing preys and predators, evolved on the basis of the robots' success in, respectively, evading or approaching one another (e.g. time to contact, [5, p.8, p.10]). Thus, the robots developed a variety of escape and hunt strategies, continuously creating new adaptational problems for the next generations of both predators and prey. Although these strategies do not necessarily become more powerful over generations, Nolfi [10, p.176] notes that there are indications that "co-evolution may produce solutions to problems that evolution of a single population cannot solve". In all then, letting robots co-evolve on the basis of their interaction with one another leads to interesting and sometimes powerful new forms of behavior.

5. Autonomy: the technical and the hard problem

On the basis of the work in evolutionary robotics, examined so far, it is possible to make the following observations. First of all, within co-evolutionary robotics, the amount of tinkering by the designer has been reduced in comparison with evolutionary robotics, and even more so in relation to cases where robots are being explicitly designed. Therefore, it is fair to say that the behavior of co-evolved robots have a higher degree of autonomy than their evolved or designed counterparts. Particularly, letting the specific details of the task to be solved be set by the evolutionary process itself (i.e. by the co-evolving opponent) is an important step towards cutting the ties that bind the robot to its designer.

However, although the amount of supervision and intervention may have been reduced in co-evolutionary robotics, it certainly has not been removed completely. It is still the human investigator who (pre-)determines the genotype-to-phenotype mapping, the mutation and crossover rate, the number and lifetime of individuals, their relation (competition, cooperation or mere co-existence) and their environment. Often, many of these aspects are tested beforehand through software simulations to determine the best experimental conditions.

Although this qualification is an important one, there is no principled reason to think that the role of the designer cannot be further reduced. In a sense, the perspective of co-evolutionary robotics transforms the issue of autonomy into a *technical problem* of finding ways to computerize all relevant aspects of the artificial evolutionary process. Though this undoubtedly presents a daunting task that is not going to be solved soon, there seems to be no principled reason for claiming that this technical problem is unsolvable.

However, many people would feel that this technical problem of autonomy does not address the real issue at all. Even if the preparation of the evolutionary process can be kept to an absolute minimum, and even if the evolutionary process can then be left entirely to its own, in what sense could the resulting robots be said to pursue goals of their own? This I take to be the *hard problem* of autonomy, in parallel to the 'hard problem' of consciousness. The hard problem of autonomy

involves the question to what extent the goals of the robots are genuinely *theirs*.

A different formulation that emphasizes the same point is given by Bourgine & Varela [1, p.xi], when they say that autonomy of systems "refers to their basic and fundamental capacity to *be*, to assert their existence and to bring forth a world that is significant and pertinent without being pre-digested in advance. Thus, the autonomy of the living is understood here both in regard to its actions and to the way it shapes a world into significance."

Autonomy is related to the fact that my goals matter to *me*, and that the events in the world and my actions are significant in relation to my goals. (It bears emphasis that this significance need not be consciously experienced, as I briefly indicated in section 2. above). Given this interpretation of autonomy, the technical progress discussed above does not even seem to touch on the subject. At least, it is hard to see how technical progress of the kind discussed above would help in solving the hard problem.

6. Autonomy and the body

So, what would help to make progress in relation to the hard problem of autonomy? A consideration of the body of systems might offer some guidance in answering this question. I suggest that it is the *embodiment* of a system that provides a fundamental ground for its autonomy. To a significant extent, it is the body, and the ongoing attempt to maintain its stability, that makes goals belong to the system. Ultimately, autonomy is grounded in the formation of action patterns that result in the self-sustenance of the embodied system. If this is correct, one may say that autonomy *develops* during the embodied interaction of a system with its environment. It grows during the performance of movements and actions and the reception of their stabilizing or unbalancing effects. In a slightly different context Sheets-Johnstone [14, p.137, p.232; see also 7] similarly emphasizes this process of development when she says: "Movement forms the I that moves before the I that moves forms movement".

However, simply connecting a robot body to a neural network that learns about the contingencies of its behavior in its environment would not lead to the development of an autonomous system, because the network and the body are not tightly coupled and do not form a fully *integrated whole*. Generally speaking, the robot's body and its control systems are too artificially connected. The robot is constructed (or bought) separately, and its specific neural controller is inserted only at a later stage. Body and controller have not evolved or developed together. Indeed, in many cases in evolutionary robotics, the control system is evolved and/or trained separately through simulations, and only later implemented in a real, physical, robot. It is well known that this can lead to substantial differences in the conduct that is displayed: "Individuals evolved in simulation do not behave the same way when downloaded into the real robots". [5, p.14]. But even apart from such behavior-related surprises, this type of 'mind-transplantation' (i.e. inserting a fully developed control system into a ready-made body) can hardly be expected to allow autonomy to develop within the system, because it does not allow any self-organization of the control-body interaction. When claims are made about self-organization or emergence in relation to robots, these claims in general refer to the process of interaction between the robots and their environment, and not to the interaction between the robot-bodies and their control systems.

As Chiel & Beer [2] have pointed out, however, it is a mistake to consider the brain or the body in isolation, as both have developed in constant conjunction during their evolutionary and lifetime interaction with the environment. The search, then, is for an approach that allows for a tight coupling between bodies and control-systems both phylogenetically and ontogenetically. For this reason, a second type of co-evolutionary robotics is of considerable interest.

7. Body, plasticity and evolution

At the moment, robotics research into the co-evolution of morphology and control-systems is just beginning [12, p.33], although there have been some fascinating early examples in the context of the evolution of virtual creatures (e.g. [15]). The main features of morphology that are studied involve the shape and position of sensors and motors, specific features of the materials that constitute the robot, and general characteristics of the robot's body such as size and weight. Due to the well-known credit assignment problem (i.e. the difficulty of determining which parts of the system are responsible for a specific behavioral success or failure), these characteristics are generally not varied all together at once, but rather in isolation from one another.

Despite its preliminary status, Nolfi & Floreano [12 p. 33-34] state that this research has led to observations of surprising robot strategies for solving specific tasks. Also, several interesting correlations have been observed between morphological characteristics (e.g. body size and wheel distance), as well as between

morphological and environmental features (in obstacle-rich environments the robots that develop tend to be smaller than in environments with few or without objects). Moreover, investigations are underway in which evolution and learning are mixed. Specifically, there are indications that the evolution of 'plastic' individuals, capable of adapting to changing bodies and/or environments during their lifetime, can lead to the production of more effective behaviors and the facilitation of scaling up behavior to more complex problems.

Thus, it seems clear that the field of co-evolutionary robotics is starting to head in the direction that is of great relevance to the topic of autonomy: the co-development of body and mind, phylogenetically and ontogenetically. Although the present status of this type of research does not warrant any firm conclusions about the autonomy of the robots thus developed, I think that the suggestion that a certain degree of autonomy is present in them is hard to avoid. If robots develop their capacity for interaction with the environment (including other agents), while learning about what their bodies and their environment afford, both on an evolutionary timescale and during their lifetime (or, if one prefers, their time of operation), what reasons could one then give to insist that the robot's goals are not genuinely theirs?

I will consider one argument that is related to the suggestion that the body is pivotal in the development of autonomy. One could insist that the bodies of robots are too artificial to allow any autonomy to emerge. Indeed, it is undeniable that the bodies of robots are fundamentally different in kind compared to the type of bodies belonging to agents that we normally grant a certain degree of autonomy: living organisms. The question is whether this difference is relevant to the issue of autonomy. In trying to clarify what makes the matter of living organisms so special, a concept that gets mentioned often is '*autopoiesis*' [8], referring to the self–generating and self–maintaining capacity of the basic building blocks of organic bodies: cells. Autopoiesis is thought to be closely connected to the way living bodies self-organize, and to the maintenance of a dynamic coupling between the system and its environment.

However, the notion of autopoiesis does not reflect some intrinsic quality of a specific kind of matter but rather indicates a characteristic of the *organization* of matter. As Maturana & Varela, [8, p.51; see also 17, p.732] say: "the phenomena they generate in functioning as autopoietic unities depend on their organization and the way this organization comes about, and not on the physical nature of their components." Given that it is the organization of the components and not their material constitution, the question is open whether autopoiesis could be realized in artificial matter. In all then, I do not think that autopoiesis provides a *principled* ground for denying any degree of autonomy to the type of robots discussed above.

The argument does indicate a further constraint on robotics however. Currently robots are constructed mainly out of metals and plastics. It is highly questionable in actual *practice* whether these types of materials allow for a genuinely autopoietic organization (see also [6]). An investigation of how different types of material would affect the behavior of the robots involved would therefore be most relevant to deepening our understanding of the relation between co-evolved robots and the concept of autonomy.

8. Conclusion

When robots are "let free to act", as Nolfi & Floreano [12, p.31] put it, and they start moving around in their arenas, are they acting autonomously? As I have tried to indicate in this paper, a consideration of the robots' embodiment provides an interesting perspective on this question. Specifically, the dynamic coupling between the body and its control system during the phylogenetic and ontogenetic development of interaction with the environment may be essential to the emergence of autonomy in the robot. Autopoiesis may be an important factor in the establishment of such a dynamic coupling, further reinforcing the importance of the body in relation to autonomy.

Of course, it is always possible to insist that robots, whatever their evolutionary and learning history may be, and regardless their material constitution, can never pursue goals that are truly theirs because they are generated through a process that is controlled by someone else. To this argument I would like to offer the following counter: When I am very hungry and I am trying to get some food, I am sure that I am trying to achieve a goal of my own that is of overriding significance to me. It would be puzzling if not annoying if someone would suggest to me that I am not really *pursuing my* goal at all, but that I am just performing a set of actions along the lines set out by an evolutionary process totally beyond my control in order to achieve purposes that are relevant to me only in virtue of that same evolutionary process. At some point, the question about whether goals are really mine stops making sense. I suggest that the body provides that point.

If my claim about the relevance of the body to autonomy is not incorrect, there seems to be no principled reason to suggest that robots could never be autonomous to a significant degree. This conclusion, however, should not blind us to the currently enormous differences between robotic and organic embodiment. Co-evolutionary robotics may be on the right track, but it surely has a long way to go.

Acknowledgments

I would like to thank the Brazilian funding agency FAPESP for financial support, the university of UNESP, Marilia, SP, Brazil for providing great conditions to work, the NICI, Nijmegen, The Netherlands for permission to work at UNESP, and Maria Eunice Quilici Gonzalez and Mariana Broens for their support during the preparation of this paper.

10. References

[1] Bourgine, P. & Varela, F.J., Towards a practice of autonomous systems, in F.J.Varela & P. Bourgine (Eds.) *Towards a practice of autonomous systems*, Cambridge, MIT-Press). 1992, p.xi,

[2] Chiel, H.J. & Beer, R.D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends In Neurociences, 20*(12), 553- 557.

[3] Clark, A. (2001). *Mindware: an introduction to the philosophy of cognitive science.* Oxford: Oxford University Press.

[4] Dennett, 1996, *Kinds of minds: Towards an understanding of consciousness.* New York: Basic Books.

[5] Floreano, D., Nolfi, S., & Mondada, F. (2001) Co-Evolution and Ontogenetic Change in Competing Robots. In M. Patel, V. Honavar, and K. Balakrishnan (eds.), *Advances in the Evolutionary Synthesis of Intelligent Agents*, Cambridge (MA): MIT Press.

[6] Haselager, W.F.G. (2003). Form, function and the matter of experience. *SEED (3)*,3, 100-111. http://www.library.utoronto.ca/see/pages/SEED%20journal%20library.html#3_3

[7] Haselager, W.F.G. & Gonzalez, M.E.Q. (2003). A identidade pessoal e a teoria da cognição situada e incorporada. In M. Broens, C. Milidoni, (Eds.). *Sujeito e identidade pessoal: Estudos de filosofia da mente.* São Paulo: Cultura Acadêmica. 75-88.

[8] Maturana, H.R. & Varela, F.J. (1987). *The tree of knowledge: The biological roots of human understanding.* Boston: Shambhala.

[9] Nolfi S. (1997). Evolving non-trivial behaviors on real robots: a garbage collecting robot. *Robotics and Autonomous Systems*, 22: 187-198,

[10] Nolfi, S. (1998). Evolutionary robotics: Exploiting the full power of self-organization. *Connection Science, Vol.10*, Nos 3&4, 167-184.

[11] Nolfi, S. & Floreano, D. (2000) *Evolutionary robotics: The biology, intelligence and technology of self-organizing machines.* Cambridge, MA: MIT Press.

[12] Nolfi, S. & Floreano, D. (2002). Synthesis of autonomous robots through evolution. *Trends in cognitive sciences, Vol.8,* No1. 31-37.

[13] Pfeifer, R. & Scheier, C. (1999). *Understanding intelligence.* Cambridge, MA: MIT Press.

[14] Sheets-Johnstone, M. (1999). *The primacy of movement.* Amsterdam: John Benjamin Publishing Company.

[15] Sims, K. (1994). Evolving virtual creatures. K.Sims, *Siggraph '94 Proceedings, July 1994*, pp.15-22.

[16] Turing, A.M. (1950) Computing machinery and intelligence.
Mind, Vol. LIX, No.236. (pp.433-460)

[17] Ziemke, T. & Sharkey, N.E. (2001). A stroll through the worlds of robots and animals: Applying Jakob van Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica, 134*- 1/4, 701-746.