The Centrifugal Development of Artificial Agents: a research agenda

Ana Sofia Esteves and Luís Miguel Botelho[1]
[1.] Supervisor
We, the Body and the Mind Research Lab, ADETTI/ISCTE
Av. das Forças Armadas 1649-026 Lisboa, Portugal
E-mail: sofia.esteves@we-b-mind.org, WWW: http://we-b-mind.org

Keywords: Autopoiesis, Adaptive Systems, Artificial Life, Autonomous Agents

Abstract

This paper presents a research agenda underlying a PhD proposal on embodied software agents. The research aims to demonstrate that it is possible, for embodied software agents, to develop first-person meanings for environmental perturbations. According to the proposed approach, first-person meanings will be created through the agent centrifugal development. We argue that centrifugally developed agents, which create first-person meanings for environmental perturbations, will perform better in identified classes of tasks. They will also contribute to advance the state of the art regarding embodiment, particularly with respect to the grounding problem. For traditional AI, the mind is a symbol processing system that can exist without the body or act in an independent way. Although traditional approaches have successfully solved several problems, they are trapped in the grounding problem. Embodied cognition approach departs from the dominant paradigm by focusing on cognition as an embodied situated activity. Organismic embodiment supporters argue that the systems development method (centrifugally versus centripetally) is decisive to the system autonomy and, thereby, to their cognitive abilities. This research will use models of cellular division and evolution from cellular biology and will propose processes, through which meaning grounding can be achieved as an emergent property of centrifugally developed agents.

1. INTRODUCTION

Artificial Intelligent systems manipulate symbols that are meaningless to them (do not have first-person meaning). This is the Symbol Grounding Problem [5, 13].

Several authors have pointed out that the grounding problem resolution will lead to better and more autonomous and intelligent systems [2, 5, 6, 13, 17].

There are two approaches to grounding: the Cognitivist Grounding and the Enactive Grounding. However, none has solved the grounding problem. In both approaches, there is a meaningless (non-grounded) mechanism that associates the symbols to their internal representations. Thus, the meanings the system builds for the symbols are not grounded.

The grounding problem can be avoided if the grounding mechanism itself has first-person meaning to the system, that is, if it is created by the system. Thus, intelligent agents must be self-developed (i.e., centrifugally developed), which is possible only if they are the outcome of co-evolution and mutual specification of the self-organized system and its environment.

This is the goal of *organismic embodiment*. Some proposals attempted to achieve this view but none has been successful.

This paper presents a research proposal, still in an early stage, and will address some of the challenges raised by this vision, specifically considering the development of software agents.

Indeed, the proposed research intends to demonstrate that the resolution of the grounding problem leads to better and more autonomous agents, in accordance with a set of criteria that will be defined; that through centrifugal development, it is possible to advance the resolution of the grounding problem; and that it is feasible to achieve centrifugally developed artificial agents.

More specifically, it aims to conceive a model inspired in biological processes that will lead to the centrifugal development of artificial agents. The biological inspiration is centered in the development of multi-cellular organisms, particularly cellular division, cellular communication and cooperation, cellular differentiation and cellular death processes.

The proposed research will cover the following five tasks. First, to identify the emergent properties of the system, or, in other words, properties of the system that can be observed but that were not deliberately planned in design-time. Second, to identify and formalize concrete, although simplified, application that can be addressed through the developed model. Third, specifying the evaluation criteria. Fourth, to apply the designed model to a sub-set of the identified problems and evaluate the extent to which the research main goal will be achieved. Finally, assess the robustness and domain independence of the projected approach.

Section 2 presents a review of the related work that has been the source of inspiration of the planned study.

Section 3 describes the proposed research. Since we are in an early stage of the program, the presented description is not a formal one. Section 4 proposes a demonstration scenario.

Section 5 summarizes the main guidelines for future developments.

2. RELATED WORK

The foundation hypothesis of traditional Artificial Intelligence (AI) is the established postulate of classic cognitive science concerning the mind-body dichotomy – the material independence postulate [11]. For traditional AI, the mind is reasonably independent of the body [10]. Cognitive science and AI main approaches – symbolic and connectionist – share this postulate [11]. AI mainstream approaches have been successful in solving a large set of problems [7]. Nevertheless both are still trapped in the grounding problem [7, 17].

2.1. The Symbol Grounding Problem

Searl [13] and Harnad [5] concluded that current AI systems are not and cannot be intelligent. The reason is that cognition cannot just be symbol manipulation. They argue that AI systems deal with symbols that are meaningless to them. Systems do not have first-person meaning for symbols. Symbols are only meaningful to the external observer or designer. Harnad coined this statement as the Symbol Grounding Problem. For the authors, cognition, and consequently intelligence, depends on the overcoming of this problem.

They illustrate the argument through two metaphors. First, Searle [13] introduces the Chinese Room argument. Locked in a room and without any knowledge of Chinese, you are given a set of instructions on your native language and a group of Chinese symbols. The instructions command you to give back certain Chinese symbols in response to the given group of Chinese symbols. From the point of view of somebody outside the room, if you correctly follow the instructions, your answers are identical to those of native Chinese speakers. The observer could say that you decided what to answer. But, in your point of view, you only manipulated meaningless symbols (they did not have first-person meaning to you). You simply behaved like a computer. Then, Harnad [5] asks the reader to suppose that he has to learn Chinese as a first language, with only a Chinese-Chinese dictionary. With this restriction, the reader could never ground the meaning of the Chinese symbols. For Harnad, this task is an impossible one. Symbols are part of a formal convention, normally providing no indication about their corresponding meaning.

Several authors have pointed that the grounding problem is not limited to symbolic representations [17]. Savage [12] concludes: "*Broadly, the grounding problem is the task of ensuring that an artificial agent's various transactions with the world have intrinsic meaning for that agent rather than resulting from assumptions and predispositions which are handed down by the agent's designer.*"

Interpreting the Harnad work, Taddeo and Floridi [14] argue that no semantic resources should be presupposed as already pre-given to the system or uploaded from an external "semantically-proficient" entity. However, the system still can have its own capacities and resources to ground symbols. The authors conclude that these limitations point a mandatory requirement for any grounding approach – the zero semantical commitment condition.

The question that arises is how to develop a system that can have first-person meaning. Some have tried to give a successful answer. However, in the next sub-section it will be acknowledged that this question is still an open one.

2.2. Approaches

Agreeing with the above arguments, many AI researchers view the Grounding Problem as a central challenge. Several main approaches can be distinguished. According to Zimke [17], agent-environment interaction is the foundation of cognition and intelligent behavior and, furthermore, grounding requires direct causal connections between the system and its environment, meaning, without being mediated by an external observer. The question that divides these approaches is how to achieve that. Still, none of the approaches solve the Grounding Problem, as it will be discussed in this sub-section.

Taddeo and Floridi [14] analyzed eight strategies and divided them in three main approaches: *Representationalism*, *Semi-Representationalism* and *Non-Representationalism*. They concluded that all strategies are semantically committed not solving the Symbol Grounding Problem. For the *Representationalist* approach, the meanings of the symbols used by a system are the representations elaborated by that system. It proposes to solve the Symbol Grounding Problem by linking the perceptual data to available representations. The main problem with the Representationalist approach is that the elaboration of the representations relies in pre-given ungrounded capacities and resources (primitive representations). The *Semi-representationalist* approach is very similar to the Representationalist approach. The difference is that the representations are elaborated through a designed learning process, which depends on pre-given or externally acquired resources. In the *Non-representationalist* approach, symbolic representations are not fundamental to achieve an intelligent behavior. Therefore, it avoids any elaboration of explicit representation. However, Taddeo and Floridi [14] argue that the Symbol Grounding Problem is delayed, not avoided. The system does not deal with the Symbol Grounding Problem initially, but, to develop some higher cognitive capacities, it will need to manipulate some symbols.

Zimke [17] divides the strategies to grounding approaches into Cognitivist grounding (*Representationalism* and *Semi-representationalism*) and Enactive grounding (*Non representationalism*). He also concludes that none has solved the Symbol Grounding Problem. In the Cognitivist approach, the meaning of the symbols used by a system consists of explicit manipulable, internal representations; and the system is composed by a central sub-system and by perceptual/transducing input sub-systems. So, the system grounds the processed symbols by creating a new internal representation, which can be the result of transducing processes or a composition of existing representations (either primitive or composed). In the Enactive approach, there is no central control and, as mentioned before, there is no need for representations. The meaning of the symbols is grounded, not only in perceptual experience, but also in action; Sensory-motor experience replaces perceptions [2].

In addition to the previously presented arguments, in both approaches, the overall system is a result of external design. So, the systems use a meaningless (not grounded) meta-control mechanism to learn the control mechanism (that is, to associate the symbols to an internal representation or action). Thus being, the meanings that the system builds for the symbols are not first-person meanings, and so, they are also not grounded.

2.3. Advantages of Grounding

Agent Based Computing is one of the most important research areas in the AI field. Wooldridge and Jennings [16] propose a weak and a strong notion of an agent. For both, an agent is a computer system that exhibits autonomy (have some kind of control), social ability (can interact with other agents), reactivity (can perceive their environment and respond to changes), and, finally, pro-activeness (can pursuit its own agenda, taking the initiative). The stronger notion – popular among those working in AI – also demands that the agent should either be conceptualized or implemented, inspired in some characteristics of animals, usually humans.

Surprisingly, as Zimke [17] states, the impact of the lack of grounding on achieving agents, increasingly more intelligent and autonomous, has been neglected. Today, agent based computing is applied to a vast collection of problems and the opportunities for new applications are always growing. One major challenge (constrain) for the agent designers is that the (agent) world is often unpredictable. Therefore, the relationship between the symbols and the external events is likely to change. So, MacDorman [6] concludes that the task of pre establishing links between the symbols and external references is insufficient and ineffective. The agent cannot adapt itself for real world unpredictability, what compromises the agent autonomy and performance. So, if the agent world is complex, grounding is a must for higher levels of intelligence, autonomy and performance.

2.4. Clues to overcome the Grounding Problem

As previously said, none of the approaches to grounding, despite their scientific value, has been able to offer a process by which the symbols manipulated by an agent acquire meaning to the agent itself. They all use non-grounded mechanisms.

In order to avoid a non-grounded mechanism to control the way the acquired symbols are grounded, we need agents that are linked in their environments, in the way organisms are. This means that we need to get inspiration in systems, which, as a whole, have emerged in interaction with their environment. For Zimke [17] the basis of intelligent behavior and meaningful interaction requires that systems and their environment must *coevolve* and become mutually dependent, forming a systemic whole. Botelho and Figueiredo [2] conclude: "*the agent architecture must use evolutionary processes through which the agent self-organizes, creates its own parts, and learns its behavioral rules in co-evolution with the environment.*" This is the goal of *organismic embodiment*, which envisions artificial agents that can construct and organize themselves as well as their own environmental embedding. The roots of this vision are the von Uexküll works [15] about centrifugally developed systems and the work of Maturana and Varela [8] regarding autopoietic systems.

Von Uexküll [15] argued that the difference between machines and organisms is their degree of autonomy, which results from the way they are developed: Machines are centripetally developed (assembling externally produced parts) whereas organisms, as "living plans", develop themselves centrifugally (the whole comes first and the parts grow from it). Usually, an agent is designed and implemented through a process of decomposition and assembling, or, using the words of Von Uexküll [15], the agent is centripetally developed. In design time, it is specified how the agent can be composed of a set of externally created components and how these components should be assembled and should interact [16]. In the implementation phase, the components are externally built and assembled to form the agent. This development methodology significantly differs from the one that organisms follow (i.e. centrifugally developed). For example, once a sperm fertilizes an egg cell, the resultant cell begins to divide itself and, through a complex dynamic process, may originate a newborn mammal. There are proposals attempting to design centrifugally developed systems, namely in the robotic field. For the moment, none has accomplished von Uexküll vision [19].

Similarly [18], Maturana and Varela [8] have distinguished allopoietic systems from autopoietic systems, stating that living systems are autopoietic systems in the physical world. In the words of Maturana and Varela [8], an autopoietic system is "*a system organized (defined as a unity) as a network of processes of production, transformation and destruction of components that produces the components which: (i) through their interactions and transformations regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it as a concrete unity in the space in which they exist by specifying the topological domain of its realization as such a network*".

In other words, an autopoietic system is a self-developed system and a special type of homeostatic systems, for which the critical variable to be hold constant is its own organization. Moreover, these systems are interactively open to their environment [10]. In order to preserve their organization they may need to adapt their structure. Maturana and Varela distinguished between the organization of a system and its structure. By organization, they mean "*those relations that must exist among the components of a system for it to be a member of a specific class*" [8]. By structure, they mean "*the components and relations that* *actually constitute a particular unity, and make its organization real*" [8]. A simplistic example: imagine a company that must be organized in a hierarchical way (e.g. The Director commands the Department Manager, and then the Department Manager commands the Programmer). In a given moment, the structure of the company is the concretization of its organization (e.g. who is in each position, how they give the commands – face to face, by fax, etc.). A change in the person that was the Department Manager does not necessarily lead to a change in the organization.

Some proposals, have attempted to achieve autopoietic systems, but none has accomplished the requirements defined by Maturana and Varela [3, 9, 19]. Most of the proposed models are simplified simulations of chemical processes in a discrete bi-dimensional space. Each position in the space is either empty or occupied by a single *particle*. Particles generally move in random walks in space. There are different particle types engaged in distinct chemical reactions. McMurrin [9] analyzed some of these models (namely, the one proposed by Varela). The author states that, in the presence of other systems, none of the proposed systems tend to maintain their organization (i.e. they merge).

Briefly, grounding may be achieved through the *organismic embodiment* vision, which requires artificial agents that are self-developed, open and self-organized systems.

Self-developed systems are able to construct themselves (its own components and their relations), and adapt their structure by generating, changing or removing its own components.

Open-systems have "perturbatory" channels between with their environment, through which they interact with and influence their environment, and vice-versa.

Self-organized systems that struggle to preserve their organization and present the five fundamental properties proposed by Correia [4], specifically i) *no external control* - the organization of the system should not be externally specified; ii) *increase in order* - overall increase of the system's order (e.g. entropy); iii) *adaptability* - must be robust against perturbations, adapting itself to changes; iv) *interaction* - the resulting behavior of the system must involve a correlation of actions to produce an organized behavior, under some criterion; and v) *asynchronism* - there is no form of global synchronization.

2.5. Conclusion

The grounding approaches agree [17] that the agent-environment interaction is the foundation of cognition and intelligent behavior. Although giving important contributions for future developments, all strategies for solving the grounding problem are semantically committed, not solving the problem because all of them use non-grounded control mechanisms.

To avoid non-grounded mechanisms it is required the co-evolution and mutual specification of the self-organized system and its environment. This is the goal of *organismic embodiment* [18], which envisions artificial agents that can construct and self-organize themselves and their own environmental embedding.

There have been some proposals to achieve the *organismic embodiment,* but none has accomplished it.

The research proposed in this paper will study and define processes inspired in cellular biology [1] and in self organized systems theory, through which, artificial agents may be self-developed systems, self-organized systems and open systems. It will also consider work regarding dynamic meaning development, upon which it will propose a conceptual and computational model capturing the relation between grounding and *organismic embodiment*.

3. PROPOSAL

The development of a multi-cellular organism is initiated with a single cell – zygote – that contains all the information about itself and about the organism that will potentially originate. This information includes capacities, functionalities and regulation mechanisms. In the first phase, this cell divides itself successively, originating many cells that have the ability to divide. In a second phase begins the cellular differentiation process of cells obtained in the previous phase. This process runs in parallel with the cell division. The cellular differentiation process is responsible for the activation or deactivation of cell specific capacities, functionalities and regulation mechanisms (depending on gene activation). Through this process, cells acquire a specific function type. Specialized groups of cells are formed and the organism tissues emerge. Organized auto-regulation processes come out from the cooperation and the coordination of the organism's components [1].

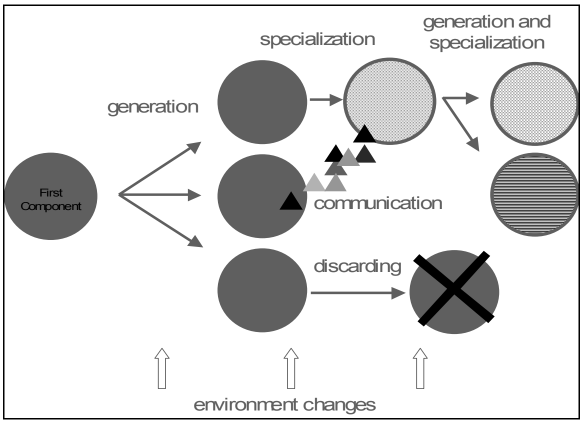

Figure 1. The centrifugal development of an agent.

The proposed research will suggest a model capable of centrifugally develop an artificial agent through a process similar to the described multi-cellular organisms development process (Figure 1).

This self-development process is initiated with a single component with limited resources and reliability. In certain environmental or internal circumstances, this original component will create new similar components, which will also have limited resources and reliability. Some of those components can also generate others. In some cases it is possible that this ability becomes inactive. At same time, some of the generated components may become specialized in certain tasks. The components can then acquire a specific type, and specialized groups of components may appear. It is also possible that some components might be discarded, either through self-destruction or due to an external cause. It may happen for a component to communicate with its neighbors, with all components, and even with the external environment.

These processes lead to the centrifugal development of the artificial agent and will continue until the agent is capable of facing the challenges of the environment and manages to satisfy its original goals.

It is important to state that there is no global coordination, no global clock and no centralized control. Through the local cooperation and coordination of the agent's components, through their continuous generation and specialization, and through the process by which some components are discarded, a system will self develop and will self-organize in a way that it will enjoy the expected properties of an agent. When an environment perturbation occurs, the agent components able to sense its impact, initiate a chain reaction that it is expected to result in the adaptation of the agent to the new environment state. It is essential to refer that this adaptation process is dynamic, since each interaction with the environment can also be seen has an environment perturbation

It is proposed that each component of the centrifugally developed artificial agent (namely the first one) should have receptors, actuators, an internal mechanism and an internal state, which includes the inherited capabilities.

Receptors are sensors to receive external signals and convert them into internal signals. There are different kinds of receptors. Each receptor can be accessible or inaccessible. Actuators are channels that enable the component to communicate with its external environment. Different types of signals must be transmitted through different kinds of actuators. Each actuator can be accessible or inaccessible. Both, receptors and actuators are perturbatory channels.The internal state is the reflection of past interactions, since it can change due to internal and external causes. The component response will depend of its internal state and, in consequence, of previews interactions. The internal state includes the inherited capabilities. The inherited capabilities are owned and transmitted by the parent component, which were given either by the designer or by a program. Each capability can be accessible (available) or inaccessible (unavailable). One example of inherited capabilities is the capability to generate other components. The internal mechanism represents all the internal processes that can happen inside the component. Those processes may be triggered by an external signal or by an internal event. An example is the process that leads to the component specialization by making inherited capabilities, receptors and actuators accessible or inaccessible.

4. DEMONSTRATION SCENARIO

The proposed approach will be applied to one demonstration scenario where the centrifugally developed agent should be able to sense a room environment and to control the music that is selected at each moment as well as its volume, trying to achieve the best room ambiance for the current context.

The room will be equipped with two multimedia equipments (a Hi-Fi system and a Projector), a media repository, a light sensor and a changeable set of furniture and objects.

The agent will be able to sense three kinds of contextual factors: physical environmental factors, social and personal factors. Physical environmental factors constitute the physical characteristics of the environment, such as light intensity. Social factors describe the social context in which the media contents will be deployed, e.g. if it is a meeting or leisure time. Personal factors describe the affective state of the involved people, e.g. if a person is tired.

The quality of the room ambiance should be a function of the individual satisfaction of the people that is present in the room.

5. RESEARCH PLAN

To reach the proposed goals we will do detailed revision of the most relevant literature. This will enable us to establish a connection among the most recent approaches from different scientific fields, expecting to reach an interdisciplinary state of the art concerning the cognition and the "grounding problem". This review will address several fields, namely cellular and molecular biology, cognitive biology, computational biology, cognitive science, artificial intelligence, psychology and the neuroscience.

After that, we will define a set of evaluation criteria that will allow us to assess the results of the proposed model, namely the system capacity to pursue its objectives and its capability to overcome the "grounding problem".

Next, we will proceed to model the system's first component. This will involve the definition of its capacities, form and type, specifically in what concerns the functionalities, regulation mechanisms and objectives. After the conclusion of the first component model we will proceed to its implementation in the proposed demonstration scenario. Then, we will start to demonstrate the concept. First, it will be necessary to prove that the artificial agent is a self-organized system. Then, the impact of different environmental perturbations on the system organization will be studied.

Through the defined set of evaluation criteria, it will be possible to evaluate the obtained results and, therefore, to estimate how far we are from reaching our main goal. The analysis of the results will also allow the identification of emergent properties of the proposed model. Then, we will assess the robustness and domain independence of the proposed approach, to accomplish our last goal.

AKNOWLEGMENTS

This paper reports work partially supported by Fundação para a Ciência e a Tecnologia through the Research Unit number U605 and by the PhD scholarship nº SFRH/BD/28287/2006.

References

[1] Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. and Watson. 2002. Molecular Biology of the Cell. Garland Publishing, Inc, New York.

[2] Botelho, Luís Miguel, and Pedro Figueiredo. 2004. "What your body and your living room tell my agent." In Proceedings of the AMAS 2004 Workshop “Balanced Perception and Action in Embodied Conversational Agents”.

[3] Chrisley, R. 2003. "Embodied Artificial Intelligence.", Artificial Intelligence, 149, no. 1: 131-150.

[4] Correia, Luís. 2006. “Self-organisation: a case for embodiment.” Technical report, University of Lisbon and AI Lab - University of Zürich., LabMAg.

[5] Harnad, Stevan. 1990. "The Symbol Grounding Problem." Physica D, 42: 335-346.

[6] MacDorman, Karl F. 1999. "Grounding symbols through sensorimotor integration." Journal of the Robotics Society of Japan, 17, no. 1.

[7] Machado, J., and L. M. and Botelho. 2006. "Software agents that learn through observation.”In Proceedings of the International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2006). Hakodate.

[8] Maturana, H., and F. and Varela. 1987. The Tree of Knowledge: The Biological Roots of Human Understanding. Shambhala, Boston, MA.

[9] McMullin, B. 2004. "30 Years of Computational Autopoiesis: A Review." Artificial Life, 10, no. 3: 277-295. MIT Press

[10] Mingers, J. 2001. "Embodying information systems: The contribution of phenomenology." Information and Organization, no. 11: 103–128. Elsevier.

[11] Ratsch, Ulrich, Michael M. Richter, and Ion. O. and Stamatescu. 1998. Intelligence and Artificial Intelligence: an Interdisciplinary Debate. Springer, New York.

[12] Savage, Tony. 2003. "The grounding of motivation in artificial animals: Indices of motivational behavior." Cognitive Systems Research, 4, no. 1 (March): 23-55.

[13] Searle, J. R. 1980. "Minds, brains and programs." Behavioral and Brain Sciences, 3:417–424.

[14] Taddeo, Mariarosaria, and Luciano Floridi. 2005. “Solving the Symbol Grounding Problem: a Critical Review of fifteen Years of Research.” IEG Research Report 21.05.05. Information Ethics Group, Oxford University and University of Bari.

[15] von Uexküll, J. 1982. "The Theory of Meaning.", Semiotica, 42, no. 1: 25-82.

[16] Wooldridge, M., and N. Jennings. 1995. "Intelligent Agents: Theory and Practice." The Knowledge Engineering Review, 10, no. 2: 115-152.

[17] Zimke, T. "Rethinking Grounding." In Understanding Representation in the Cognitive Science, A. Riegler et al., Plenum Press, New York.

[18] Ziemke, T. 2003. "What's that thing called embodiment?" In Proceedings of the 25th Annual Meeting of the Cognitive Science Society.

[19] Ziemke, T. 2004. "Embodied AI as Science: Models of Embodied Cognition, Embodied Models of Cognition, or Both?" in Embodied Artificial Intelligence, F. Iida et al., eds., Springer Press 3139: 27-36.