Towards Representational Autonomy of Agents in Artificial Environments

Argyris Arnellos, Spyros Vosinakis, Thomas Spyrou and John Darzentas

Univ. of the Aegean, Dept of Product & Systems Design Engineering, 84100, Syros, Greece
{arar, spyrosv, tsp, idarz} @ aegean.gr

Abstract. Autonomy is a crucial property of an artificial agent. The type of representational structures and the role they play in the preservation of an agent's autonomy are pointed out. A framework of self-organised Peircean semiotic processes is introduced and it is then used to demonstrate the emergence of grounded representational structures in agents interacting with their environment.

1 Autonomy and Representations in Strong Agency

There is an interesting interdependence between the three fundamental properties of *interactivity, intentionality and autonomy* which are used to describe an agent. As it is suggested in [1], there is no function without autonomy, no intentionality without function and no meaning without intentionality. The circle closes by considering meaning (representational content) as a prerequisite for the maintenance of system's autonomy during its interaction. Moreover, the notion of representation is central to almost all theories of cognition, therefore being directly and indirectly connected with fundamental problems in the design of artificial cognitive agents [2], at the pure cognitivistic framework as much as at the embodied and dynamic approaches [3]. Although an embodied agent seems to be able to handle very simple tasks with only primitive stimulus-response actions, its cognitive capabilities cannot scale to tackle more complex phenomena. These and other problems are evidences that the use of representations, even in reflexive behaviors, becomes essential [4]. However, representations should not be generic, context-free and predetermined, but they should be an emergent product of the interaction between an agent and its environment [2].

2 Emergent Representations via Self-organised Semiotic Processes

Self-organised and embodied systems admit no functional usefulness to representations. Based on the abovementioned, the incorporation of a process to support the vehicle of the representation which carries internal information about an external state seems imperative. This process should give the interactive dimension to the self-organising system and furthermore, it should correspond to the embedded structure of emergent representations. Peircean semiosis [5] can be seen as the process which drives the system into meaningful interaction. In the proposed framework, intelligence is not considered as an extra module, but as an asset emerging from the agent's functionality for interaction and the aim is the unification of the modality of interaction, perception and action with the smallest possible number of representational primitives. The present attempt is in correspondence with contemporary works in AI, such as [6] and [7]. In the present paper, there is an attempt to design a more generic architecture which will integrate aspects of self-organisation and embodiment with Peircean semiotics. There is in no way a demonstration of a totally autonomous system, but the introduced architecture overcomes the symbol-grounding problem, which is the fundamental obstacle for the frame problem, and by doing so, it introduces a type of representational structures that are integrated into the functional structure of the artificial agent.

2.1 The Structure of Peircean Signs

The basic structural element of the proposed framework is the semiotic component. A possible representation is to use a frame-like structure, and to let individual slots express the respective qualities (*qualisigns*) of the object they represent. For indexing and interpretation purposes, two more slots should be reserved to describe the unique *id* of the component and the type of data it holds. In the case of artificial environments, possible objects that can be represented in the agent's knowledge base using semiotic components are *entities*: the individual visual elements that exist as geometries in the environment. The semiotic component should possibly contain their spatial properties (e.g. translation, rotation, bounding box size) and other custom qualities that better describe their nature. Semiotic components could also describe: *relations*, i.e. spatial (e.g. near), structural (e.g. part-of) or other relations between entities, *situations*, i.e. a collection of objects and relations between them that describes (part of) the environment and *actions*, i.e. preconditions (described as the initial situation), performance (series of motor commands) and effects (changes between initial and final situation). The slots can contain either crisp values or sets. In the latter case, the component describes not just one object but a category (*legisign*).

2.2 Self-organised Peircean Semiotic Processes

The abstract architecture rising from the interaction of a self-organised system with its environment based on Peircean semiotic processes is shown in Fig. 1. A detailed analysis of the architecture is given in [8].

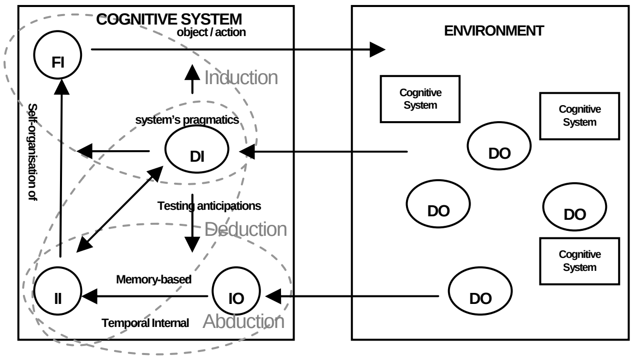

Fig. 1. An agent engaging in self-organised semiotic processes with the environment.

As a first step towards a computational methodology for implementing the proposed framework, an example has been set up, where agents are wandering around an environment and try to learn simple actions. Each agent has its own abilities concerning perception and action and initially it has no representational structures regarding possible actions. A perception mechanism, which is constantly being informed by the environment, creates the Immediate Objects (IO) as components that will drive the semiotic process. These are stored in the *short term memory*, which agents are constantly examining and comparing to their representational structure to try and detect any *surprising phenomena*, i.e. objects that they cannot categorize. In the implemented example, the semiotic components describe *entities*, *spatial relations*, *situations* and *actions*. The process of semiosis is initiated by an agents's failure to categorize an observed situation. A completed semiosis consists of the three inferential procedures: *abduction*, *deduction* and *induction*, which drive the agent's logical argumentation.

Abductive Phase

The first part of abduction consists of the observation and description of the nature of a surprising phenomenon on the basis of the anticipations of the agent. Hence, the interaction initiates from the dynamic object (DO), the environmental element of interaction. A representamen contains several IOs which in turn refer to several DOs. Which IO will eventually be actualized depends on the cognitive system's anticipations. The decision made is determined by the highest similarity score. If it reaches below a certain threshold, the IO is treated as belonging to a new category and is stored in the representational structure. A new semiotic component is created that contains all differences between the IO and the category with the most similarity. In the second part of the abduction, an analogy between the surprising phenomenon and the agent's anticipations is attempted, in order to indicate a possible direction of a hypothesis explaining the surprising phenomenon. The differences between an observed action a and a known category A can be found in both preconditions and effects of the action and may involve differences in quality values of the same entity, in the entities that take part in the action and in the relations between them. At the final part of the abduction a formulation of a possible explanation for the surprising phenomenon takes place. At this point the immediate interpretant (II) has been formed. In the end of the abductive phase, A' is created as a copy of A to describe the revised category if the hypothesis were true.

Table 1. Laws of agent's logical argumentation based on differences between an observed action a and an action category A

Difference	Hypothesis	Direction	Actions
Different Preconditions	$a \in A$	Generalize preconditions	Quality difference → expand set
			Relation difference → remove relation
			Entity difference → remove entity
	$a \notin A$	Specialize effects	Exclude quality value from set
Different Effects	$a \in A$	Generalize effects	Quality difference → expand set
			Relation difference → remove relation
	$a \notin A$	Specialize preconditions	Exclude quality value from set

Deductive Phase

In the deductive phase the consequences of the hypothesis formulated in the abductive phase are examined. In the first part of deduction, a possible direction of the consequences of the hypothesis is indicated based on the agent's anticipations. In the second part, the formulation of the consequences of the hypothesis takes place. Hence, there will be some tests needed in order this core meaning to be temporarily stabilized into a dynamic interpretant (DI). This process is the most complicated one as the self-organised system will try to incorporate the new representational structure (II) in its functional organisation. In the example, during the deductive phase the effects of the hypothesis are applied to A'. The generalization and specialization mechanisms that can take place in order to restructure an action category in the agent's representational structure, based on the type of difference and the hypothesis are summarized in Table 1. Both A and A' are kept in memory and linked to each other as A' is a descendant action of *A*. So, a surprising phenomenon will either create a new tree as a single node, or expand an existing tree by adding a descendant node to the most similar of its nodes.

Inductive Phase

The, in a way, objective meaning, which results from the semantic and pragmatic processes, should be open to revision. In case of acceptance, the hypothesis can be used to account for similar surprising phenomena in the future. Then, a new belief would be fixed and if such a hypothesis continues to persist through the agent's interaction with the environment, it will grow to a habit (FI), where a representational structure coincides with the intentionality of the respective object. In the example, if the perceived context meets the preconditions of an action, and that action involves at least one entity of type 'agent', the agent's behavior tries to imitate the action. If there is an anticipated change in the agent's position, the agent actually changes its position in order to meet the changes in the action effects. Each action in memory is assigned a score, and, whenever it is observed in the environment or it is the most similar to an action observed in the environment, its score is increased and the score of all other nodes in the same category tree is decreased. The nodes whose score is below a certain threshold are deleted from memory. With this process the agent manages to test its hypotheses by trying them, and to reinforce the correct ones, leading to the restructuring of its representational structure.

4 Conclusions and Future Work

An example has been presented as an application of the proposed framework, where agents evolve their own representational structures regarding new actions by observing the environment and trying to interact with it. The structure of the Peircean semiotic processes overcome the symbol grounding problem as they are already grounded by their nature. The fact that a representamen mediates between the DO and its interpretant provides a Peircean semiotic process with an embodied structure, since now agent's anticipations are grounded in agent-environment interaction. The authors plan to extend the analysis and implementation in more complex environments, where the representation of actions allow agents to anticipate long-term actions by embedding them seriously into time and enriching their degree of representational autonomy.

References

1. Collier, J.: Autonomy in Anticipatory Systems: Significance for Functionality, Intentionality and Meaning. In: Dubois, D. M. (eds.): The 2nd Int. Conf. on Computing Anticipatory Systems. Springer-Verlag, New York (1999)
2. Christensen, W. D., Hooker, C. A.: Representation and the Meaning of Life. In: Clapin, H., Staines, P., Slezak, P. (eds.): Representation in Mind: New Approaches to Mental Representation, Oxford: Elsevier (2004)
3. Anderson, M.: Embodied Cognition: A field guide. Artificial Intelligence. 149 (1) (2003) 91–130.
4. Pfeifer, R., Scheier, C.: Understanding Intelligence..MIT Press, Cambridge, M.A (1999)
5. Peirce, C. S.: The Essential Peirce. Selected Philosophical Writings. Vol. 1 1992 and 1998.
6. Vogt, P. (2005) The emergence of compositional structures in perceptually grounded language games.: Artificial Intelligence. 167(1-2) (2005): 206-242
7. Roy, D.: Semiotic Schemas: A Framework for Grounding Language in the Action and Perception.: Artificial Intelligence. 167(1-2) (2005): 170-205.
8. Arnellos, A., Spyrou, T. and Darzentas, J.: Towards a Framework that Models the Emergence of Meaning Structures in Purposeful Communication Environments. The 47th Annual Conf. of the Int. Society for the Systems Sciences (ISSS) 3(103) (2003).