The role of identity in agent design

(Extended Abstract)

Ines Di Loreto
ines.diloreto@fastwebnet.it

Fabien Hervouet
LIRMM/CNRS, University of Montpellier, France
fabien.hervouet@lirmm.fr

Categories and Subject Descriptors

I.2.0 [**Artificial Intelligence**]: Philosophical foundations

General Terms

Theory

Keywords

Agent, Identity, Uniqueness, Embodiment, Enaction

1. INTRODUCTION

This year AAMAS conference introduced a perspective track for "papers that analyze in some way the agent research community". The aim of this track is to understand what the trends are in agent research and foresee possible future directions. Instead of looking at where the agent community is going in an emergent way analyzing numerical trends, with this article we want to suggest where agent research could go but is not yet going. We are of the opinion that in the agent research community most of the current trends originate from the translation of particular concepts – mostly from analytic philosophy – which are only a particular western way to look at philosophy and agents. With this article we want to suggest that other paths originating from philosophy could be taken into account in order to create different directions in agent design.

2. THE ROLE OF IDENTITY: FROM PHILOSOPHY TO AGENT DESIGN

Analyzing current trends in agents design we observed that while trying to model and reproduce humans and societies, agent design mostly does not use a structured construction of the identity concept. In the rest of this paper we will support this position analyzing the identity concept, paralleling agent design and contemporary philosophical assumptions about the concepts of uniqueness, body and mind.

2.1 The concept of uniqueness

There is almost no debate about uniqueness in agent design. More generally uniqueness is in essence an issue for computer science. Any data can be copied and replicated with an absolute guarantee of ending up with two exact similar objects. This interesting property prevented researchers from really tackling such an issue. As a result a certain part of agent design seems to work on a "Universal Agent", deriving from a "Universal Man" theory from the philosophy of Plato and the republican ideals of equality, which in essence does not need a structured uniqueness to be implemented. In Multi-Agent System (MAS) identity is mainly structured from the point of view of the role of the agents. MAS usually put a multitude of agents together in order to accomplish a certain global task or to have a certain global behavior. This means that even if agents do not act exactly the same in a local way, they often originate from the same piece of code that takes into account some predefined interactions with their peers.

However, while computer science seems not to care about uniqueness, in the 1970s Maturana and Varela addressed the complex problem of autonomy, knowledge and identity in biology [9]. They characterized living organisms by coining the concept of *autopoiesis* which is defined as a complex incessant process of self-production of the system by itself, replacing its components to compensate for continuous external disturbances. In short an autopoietic system can be seen as a homeostatic system whose invariant principle is its own organization (seen as the network of relationships that defines it). Therefore in this context uniqueness may be defined as this historical coupling, i.e. the historical adaptive activity of the structure in order to fit the organization. The autopoiesis theory has inspired some scholars in the artificial life and agent design domains such as [4, 3, 13], but it remains generally too few addressed.

In the same way psychology has discussed the concept of uniqueness from its very beginning. In the '60s, Piaget led the constructivist movement promoting the vision that every individual has the ability to hold their own reconstruction of reality. This theory of knowledge supports the fact that identity is perpetually in construction, deriving from our own adaptation process in direct confrontation with the environment. This point of view will be explored a few years further by Varela et al. and their theory of *enaction* [12]. The enaction paradigm postulates the co-emergence of both cognition and perceived world through the performative body in action in the environment. Therefore uniqueness can only appear within a pure bottom-up mechanism.

2.2 The concept of body

Regarding the "body concept" it is very interesting to notice a fundamental difference in the agent design approach

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

between computer science and robotics. Indeed, while computer science focused mainly on disembodied reasoning capabilities, robotics was created with the idea of body, of physical interaction, and it is based on the experimental principles of physics and mechanics in a very grounded manner. Therefore, the advent of robotical agent design is a major step towards the consideration of the notion of embodiment for intelligent agents. Three major kinds of robotic architectures emerged : *deliberative architectures* [10] using symbolism and generally organized into multiple hierarchical layers, *purely reactive architectures* [6, 1] built by stacking finite state machines without reasoning nor symbolization, and *hybrid layered architectures* like [8] combining the advantages of behavioral and deliberative architectures.

Nevertheless, as Ziemke argues in [13], even if it has been recognized for a decade that embodiment is a necessary condition to characterize living organisms [11, 12] and that more and more researchers have attempted to address this absolute need for embodied cognition (in the developmental robotics community for instance, see [2] for a recent survey), robotics "is largely 'stuck' in the old distinction between hardware and software". Cognitivist vision is still largely dominant over a pure varelian enactive vision [13].

2.3 The concept of mind

In the 20th century problems deriving from the division between mind and body also become evident in philosophy. One of the most well-known 20th century philosophical movements is analytical philosophy. Very simply, analytical philosophy is characterized by the application of a logical method to traditional philosophical problems often using modern formal logic and language analysis. Computer science has deep analytical foundations, since the von Neumann's vision of cognition as logical problem solving.

Nowadays although the metaphor of the agent as a symbol interpreter is always present, more complex models of agents have been proposed. For instance, we can cite the Belief-Desire-Intention model [5] which is a widely used more complex model articulated around the notion of knowledge in pure bodyless approach.

Furthermore in the same pure mind-only way, interactions between artificial agents have been historically only communicational. Languages designed were nothing but logical formalized protocol philosophically based on Austin's and Searle's speech acts theories. We think this approach is inherited from the ideas of philosophers like Wittgenstein relayed later by the behaviorist psychology of Skinner. For these authors the only way we can study thought is to look at verbal behavior because, unlike in private thoughts, the behavior can be *scientifically* verified. The legacy of analytic philosophy is the vision of the mental representation.

This logical vision combined with an omniscient point of view in agent design has shown its limits for researchers who wanted to create more subjective and complete agent by-passing the mind-body dualism.

3. PERSPECTIVES

In the first part of this paper we have shown that contemporary psychology, philosophy and even biology have interesting ways of looking at the concepts of identity, uniqueness, performance and environment as interlaced and interacting. Although the problem solving vision is useful in many ways, integrating different concepts can lead to a more global vision about autonomous agent design. Obviously identity is only one of the concepts that could be analyzed and the analysis proposed in this paper makes up only a subset of the concepts that can consist in identity. For a more in depth analysis of the concept of identity in agent design see [7].

Based on our analysis we suggest that agent design can integrate the following concepts.

Uniqueness The concept of uniqueness could be very interesting to integrate in agent design in a mixed environment involving virtual agents as well as human agents in order to bond more easily with each other.

Autopoiesis The concept of autopoiesis is strictly linked with the uniqueness one. In autopoietic systems uniqueness may be considered as a particular trajectory of the coupling between organization and structure.

Enaction The concept of enaction could be integrated in agent design in order to overcome the dualism of mind/body. Going beyond this dualism can help to create agents which are more adaptive to unknown environments thanks to their deep physical grounding.

At the end of this short paper we can then say that the concepts we suggest integrating in the agent design paradigm are nothing more than necessary steps – but not necessarily sufficient – to reach the autonomy stage. However we strongly believe that as long as the design of agents is mainly based on analytic philosophy, we can only have an enlargement of the domain and not a paradigm shift which is at the basis of major advances in science.

4. REFERENCES

[1] R. C. Arkin. Motor schema based navigation for a mobile robot: An approach to programming by behavior. In *International Conference on Robotics and Automation*, volume 4, pages 264–271. IEEE, 1987.

[2] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida. Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development*, 1:12–34, 2009.

[3] R. D. Beer. Autopoiesis and cognition in the game of life. *Artificial Life*, 10:309–326, 2004.

[4] P. Bourgine and J. Stewart. Autopoiesis and cognition. *Artificial Life*, 10:327–345, 2004.

[5] M. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.

[6] R. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1):14–23, 1986.

[7] I. Di Loreto and F. Hervouet. Identity: the contribution of contemporary philosophy to agent design. Technical report, LIRMM, Jan 2012.

[8] R. J. Firby. *Adaptive Execution in Complex Dynamic Worlds*. PhD thesis, Yale University, 1989.

[9] H. R. Maturana and F. J. Varela. *Autopoiesis and Cognition: The Realization of the Living*. Reidel, 1980.

[10] N. J. Nilsson. *Principles of Artificial Intelligence*. Tioga Publishing Company, 1980.

[11] R. Pfeifer and C. Scheier. *Understanding intelligence*. MIT Press, 1999.

[12] F. Varela, E. Thompson, and E. Rosch. *The Embodied Mind: Cognitive science and human experience*. MIT Press, 1991.

[13] T. Ziemke. Are robots embodied? In *Lund University Cognitive Studies*, pages 75–83, 2001.