Embodied cognition through cultural interaction

Bart De Vylder, Bart Jansen, and Tony Belpaeme
Artificial Intelligence Lab
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels, Belgium
{bartdv,bartj,tony}@arti.vub.ac.be

Abstract

In this short paper we describe a robotic setup to study the self-organization of conceptualisation and language. What distinguishes this project from others is that we envision a robot with specific cognitive capacities, but without resorting to any pre-programmed representations or conceptualisations. The key to this all is self-organization and enculturation. We report preliminary results on learning motor behaviours through imitation, and sketch how the language plays a pivoting role in constructing world representations.

1. Introduction

In philosophy human reason and language has often been described from a functional view point, detached from any physical substrate (e.g. Fodor, 1983). By seeing the mind as a computing machine, and reason, representation and language as mere programs that run on this computing brain, they have neglected the importance of embodiment in understanding the how and why of human cognition. Likewise, language has had to endure the strangle hold of functionalism, which defends a simpli

ed view in which linguistic capabilities are the result of an innate language faculty. Learning a language and all the competences that are needed to use language, such as conceptualisation or vocalisation, is in this view merely a matter of setting the right parameters in a genetically specified program (e.g. Chomsky, 1965).

In this paper we describe work in progress, which argues against functionalism. We believe that language and cognition are not the product of an innate endowment, but rather result from the interactions between the environment and the individual and from interactions between individuals. The crux here is that language and cognition are the result of social and cultural interactions. This is easy enough to accept for language, but why should cognition find its origins in social interaction? To not bite of more than we can chew, we will limit ourselves here to conceptualisation and demonstrate how conceptualisation can be determined by social and cultural interaction.

2. Experiments in language and cognition

Some have already demonstrated on robotic platforms how embodiment plays a crucial role in acquiring representations and conceptualisations of the world (e.g. Belpaeme et al., 1998; Roy, 2000; Vogt, 2001; Steels, 2001). What is unique about the setup presented here is that the robot not only perceives the environment, but also is able to actively engage in its environment through the use of its robot arm. As such the system does not only acquire perceptual conceptualisations, but also conceptualisation on action and time. This will form the foundation for other ongoing research in endowing the robot with fully syntactic language acquisition.

The physical setup consists of a robot arm, a stereo camera mounted on a pan-tilt unit and a controlling computer. The arm has six degrees of freedom and is equipped with a gripper and its forward and inverse kinematics are known. Apart from manipulating objects, the arm enables the agent to grab and push objects and to direct attention of others by pointing to objects. The vision system consists of a stereo head mounted and a pan-tilt head, allowing the camera to look around the room and also focus on a work space with objects in front of the agent.

2.1 Imitation

We use imitative interactions to let the robot conceptualise its motor behaviour. Concretely we have defined a procedure, dubbed the imitation game, which we have used to demonstrate how conceptualisation of actions can emergence without explicitly predefining any actions for the robot (Jansen et al., 2003). Furthermore, as the imitation game is played between several agents, all agents acquire a shared repertoire of actions. This even works when the bodies of the agents are dissimilar; consider two agents with a completely different embodiment, for instance

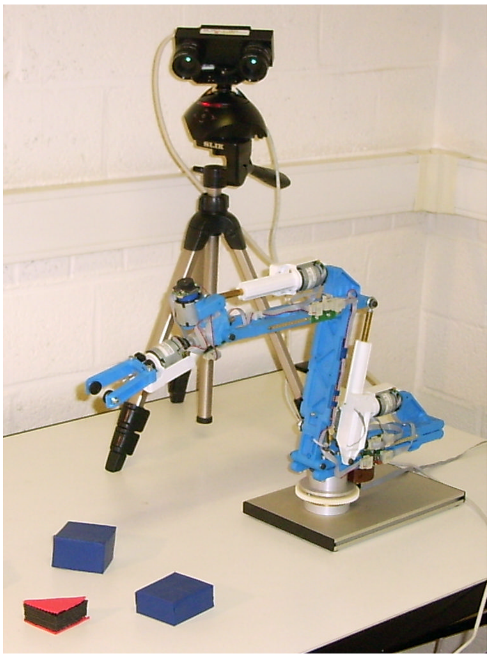

Figure 1: Robotic setup showing the 6 DOF arm and the stereo camera on a pan-tilt head.

a different number of degrees of freedom. Each agent tries to imitate the other's actions as good as possible. If one agent can not imitate the other agent's action, simply because its kinematics prohibits the performance of such an action, the imitation might fail. If this action is repeatedly not imitated successfully, the action will disappear from the agent's repertoires. As such, this leads to a shared repertoire consisting of all actions that can be observed and executed by all agents, and as the imitation game does not require the parties taking part in the game to have identical embodiments, it is possible for a robot to acquire action conceptualisations from humans.

2.2 Presence detection and shared attention

In order for an agent to be able to participate in interactions with another agent or a human, it is necessary that the participating agents can establish a form of joint attention. This enables them to narrow down the perceptual information flow and focus on the topic of conversation. For example when determining the meaning of a verbal expression, knowing what part of the physical space the speaker pays attention to, might resolve possible ambiguity left in the linguistic description. There are different ways in which joint attention can be achieved. Gaze direction, pointing, head movements, gestures etc. are all non-verbal means by which humans can direct the attention of other persons. If we want a robotic agent to participate in linguistic communication in which it has to take up the role of both hearer and speaker, it must be capable of recognising such social cues as well as providing them to other agents. One of the important social cues to direct the attention is gaze. It is therefore necessary for our robot to be able to detect faces and, if close enough, eyes and their gaze direction. Human activity is detected by a simple motion detection algorithm, while face detection is done using a multi-layer perceptron to classify skin coloured pixels. After judging possible candidates on area and shape, and fusing with depth information from the stereo vision, eye candidates are found in the faces and their gaze direction is extracted.

Acknowledgements

Bart Jansen is sponsored by a grant from the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT). Tony Belpaeme is a postdoctoral fellow of the Flemish Fund for Scientific Research (F.W.O. Vlaanderen). This research is in part funded by the COGNIRON project (FP6-IST-002020).

References

Belpaeme, T., Steels, L., and van Looveren, J. (1998). The construction and acquisition of visual categories. In Birk, A. and Demiris, Y., (Eds.), *Proceedings of the 6th European Workshop on Learning Robots*, Lecture Notes on Artificial Intelligence, Berlin. Springer.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. The MIT Press, Cambridge, MA.

Fodor, J. A. (1983). *The Modularity of Mind*. The MIT Press, Cambridge, MA.

Jansen, B., De Vylder, B., Belpaeme, T., and de Boer, B. (2003). Emerging shared action categories in robotic agents through imitation. In Dautenhahn, K. and Nehaniv, C. L., (Eds.), *Second International Symposium on Imitation in Animals and Artifacts 2003*, pages 145–152. The Society for the Study of Artificial Intelligence and the Simulation of Behavior.

Roy, D. K. (2000). Learning visually grounded words and syntax of natural spoken languages. *Evolution of communication*, 4(1):33–56.

Steels, L. (2001). Language games for autonomous robots. *IEEE Intelligent Systems*, sept-oct 2001:17–22.

Vogt, P. (2001). Evolution of grounded symbols in communicating robots. *Evolution of Communication*, 4(1):87–116.