On Learning and Representations in Cognitive Architectures for HRI

Nikolas J. Hemion
Aldebaran AI Lab
168Bis–170 Rue Raymond Losserand
75014 Paris, France

ABSTRACT

The goal of research in the field of cognitive architecture is to create *general intelligent behavior*. In contrast, the state-of-the-art in the field of Human-Robot-Interaction is to handcraft systems that are specially tailored to certain scenarios, thus creating specific rather than general solutions. In this paper I subscribe to the view that we need to understand how robots can *learn* to interact, rather than to try to hard-code general intelligent social behavior. I further describe new trends in the field on how to implement a knowledge representation for a robot and provide further ideas of how such a knowledge representation could be learned.

1. INTRODUCTION

The goal of research in the field of cognitive architecture is to create *general intelligent behavior* (10). To this end, a cognitive architecture implements a scientific hypothesis about what aspects of cognition are independent of task (8), that is to say, it is explored whether a single theory of what is common among many cognitive behaviors can support all of cognition (11). This is different from developing systems that perform particularly well in a specific task, as it is rather the goal to develop an integrated system that can cope with *many* situations.

However, state-of-the-art robotic systems in the field of Human-Robot-Interaction (HRI), like in many other robotic applications, make integral use of knowledge that the human designer has of the robots' tasks. For example, they rely on handcrafted symbolic representations, or the detection of predefined keywords. This *hardcoding* of knowledge into the robot's system on the one hand allows current robots to tackle complex problems (such as engaging in social interaction with humans) in the first place, but on the other hand renders the implementation of the robot's cognitive system *task-specific* by definition (which is incompatible with the goal of cognitive architecture to develop *general* intelligent behavior).

Cognitive Architectures for Human-Robot Interaction Workshop at HRI'14

March 3rd, 2014
Bielefeld, Germany
http://www.tech.plym.ac.uk/socce/staff/paulbaxter/cogarch4hri/

In recent years, this approach of manually engineering intelligent robots has been criticized to only work sufficiently well in scenarios where the human designer of the system can predict all possible situations in advance, and in turn can prepare the robot by providing it with handcrafted representations and algorithmic solutions. However, in cases where the environment is very complex or difficult to model, these systems tend to break down (1; 19). In response, the research field of *developmental robotics* has formed. Research in this field subscribes to the idea, that it should be tried to completely refrain from engineering the cognitive system of a robot (i.e. hard-coding knowledge into the system which thus inevitably results in a task-specific solution), and that instead the research goal should be to find a way to create robots that can *develop* and *learn* (taking the juvenile stage of biological organisms as inspiration).

2. COGNITIVE ARCHITECTURE IN DEVELOPMENTAL ROBOTICS

While many promising results have already been achieved in the field of developmental robotics, the current state of research on cognitive architecture under this paradigm is still far from an understanding of how to build robots that can develop up to a degree where they can learn to socially interact with humans.

On a methodological level, approaches to cognitive architecture that qualify as "developmental" approaches (i.e. that support learning and/or refrain from making use of hard-coded knowledge) have in common that they subscribe either to the research paradigm of *connectionism* (i.e. basing the system on the use of artificial neural networks) or to the research paradigm of *dynamicism* (i.e. modeling the cognitive system as a set of coupled dynamical systems). However, in the details of their implementation, individual cognitive architectures that have been proposed so far in the literature differ substantially from one another (e.g. 3; 6; 9; 13; 17), and no convergence to a single methodology can yet be observed. But in most of these works, the approach taken to define the system on the architectural level is to introduce a new form of structural element (above the level of complexity of neural networks or dynamical systems), as "building blocks" in the cognitive system. The overall behavior of the system is thus the result of the parallel working of such building blocks, which are simultaneously active and collaborate in processing inputs and producing outputs, without requiring any form of supervisory component or "cognitive module". But further than that there is not much agreement about

what should be the exact nature of the building blocks in a cognitive architecture.

3. REPRESENTATIONS FOR INTERACTION

To allow the robot to participate meaningfully in social interaction, the functionality of the building blocks not only needs to allow the robot to for example learn about and recognize elements of the environment, but also to learn about and to make use of recurring "patterns of interaction" that are normative in a given culture. Infants master the capability to interact according to cultural norms and rules early on in life, through playful interactions with their peers and caregivers. Interestingly, they do not need to learn a certain pattern of interaction perfectly before they can utilize it, but instead they simply try out what they have learned in an "imperfect" manner, and from these experiences improve their knowledge based on the feedback they receive. In the developmental psychology literature, the concept of such patterns of interaction is discussed under the name of *frames*: Fogel describes frames as regularly recurring patterns of communication that are each time dynamically reconstructed by the interactants (4), such as bedtime routines. Tomasello stresses that frames are defined intentionally, meaning that they establish a common ground between the interactants about what the purpose of the interaction is and thus facilitate the understanding of what the communicative intentions of the interaction partner are (18). Importantly, a coarse understanding of what constitutes a certain frame already allows the infant to participate in the interaction in a meaningful manner, even though he or she does not yet understand every detail of the interaction or all the words that the caregiver might use.

Most recently, it has been proposed by Wrede and colleagues that this capability, to first acquire a coarse understanding of an interaction and its purpose to allow a robot to participate in interactions and to receive feedback to refine its understanding, should be the basis of social learning capabilities in robots, not only for language but also for the domain of action (20). From this perspective, to learn to interact socially the cognitive architecture of a robot needs to support a frame-like representation that allows the robot to recognize patterns of interaction from the continuous stream of information in an ongoing interaction and to participate in the interaction. Previous attempts to model a frame-like representation in a computational system were making use of handcrafted symbolic representations (12; 16), which however turned out to be rather unsuccessful due to the very rigid nature of the resulting system. But the concept of a frame representation as described above understands it as very flexible and bendable in nature, which not only allows the robot to make errors but effectively *requires* it to do so, in order to elicit corrective feedback from tutors.

For the goal to develop such frame-like functionality in a system, the related concept of schema from the field of Psychology (2; 14) provides a suitable theoretical ground for the modeling of a building block in the cognitive architecture to house the knowledge of the system (6). Schemata have been proposed to be the format for the representation of knowledge in the human conceptual system, and are argued to have a frame-like structure (2). In an early neural-network model, Rumelhart and colleagues have demonstrated that frame-like representations can be implicitly stored in a distributed code in neural network weights (15). In their model, the system acquires through statistical learning the basis for frame-like functionality without being given the structure explicitly: Rumelhart et al.'s model learns a schema representation for rooms by being presented a number of descriptions of different rooms (such as, an office has a desk, a chair, a telephone, etc.). When queried with an incomplete description of a room (for example, a room with a telephone), the network's activation dynamics drives the network into a state corresponding to a coherent room representation (co-activating nodes for desk, chair, etc.).

The fundamental logic of this models provides an interesting perspective on the problem of a developmental robot having to learn frames and matching them to the current situation. From the available (incomplete) information, the system dynamically infers what other aspects of the environment and actions are likely to become relevant in the current context. This would be achieved by driving the system's state into an attractor point where additional representations become co-activated (cf. the concept of *embodied simulation* (2; 5)), corresponding to a larger-scale representation of the robot's situation. However, Rumelhart et al's model processes information in a very passive manner: Information is presented to the model, which in turn drives the system into a new state. In contrast, infants actively query their environment for information as they engage in playful interaction with their (social) environment.

This active participation, together with another important function of a schema representation, might help the infant in overcoming another difficulty in learning to interact: That of acquiring knowledge about what elements of the environment are relevant to a certain situation (which is related to the *frame-problem*). This other important function of the schema representation is that it can be used to make predictions about the environment, forming the basis for categorization (2): If an object behaves the way that a certain schema *predicted* that it would behave, then the object is categorized as belonging to the category that the schema represents (for example, a ball is recognized as a ball because it behaves in the way that we expect a ball to behave; cf. also Piaget's idea of *accommodation* (14)). In previous work we have applied this fundamental idea in a computational model, where we could demonstrate that it allows the system to self-organize the learning of several internal models simultaneously, without requiring any form of supervision of the learning process or labeling of training data (7). More specifically, we demonstrated that a robot can learn how the movement of the own body will effect the visually perceived position of the own hand, without even knowing in advance what the own hand is or what it looks like, up to achieving the same level of performance as when using labeled training data and supervised learning. This was achieved by letting the system try to learn multiple internal models at the same time, and letting these internal models compete against each other to obtain new training data: An observation is used for the training of that internal model which best predicted the observation (i.e. had the lowest prediction error).

The same learning principle could also be applied to the domain of social interaction, where a robot needs to learn

about what aspects of the environment and what actions are relevant to a certain situation, in the following way. Imagine a robot has acquired through exploration a set of actions, some of which could also involve social interaction with others: For example, in a situation where there is a person present, pointing at an object will result in the person handing that object to the robot. Once the robot has learned a set of such actions, it can make predictions about outcomes to "explore" the social environment: If the robot points at an object with the prediction that it will be handed the object, and the interaction partner does indeed hand the object to the robot, it may be concluded that the interaction partner saw the action as appropriate for the current interaction. Using this logic to "activate" relevant actions and other conceptual primitives (e.g. for objects), and combining it with a learning of co-occurrences and temporal closeness in a way comparable to how Rumelhart et al.'s model learns about the co-presence of room features, could lead us to a method for an autonomous acquisition of a frame representation.

4. SUMMARY

In this paper, I have briefly outlined the approach of developmental robotics to the topic of cognitive architecture, where it is commonly tried to model the system on the basis of some form of "building block". In subscribing to this view, I have drawn inspiration from recent work on the intersection of developmental psychology and robotics, where a bendable frame-like representation was proposed to underlie the capability for social interaction, and have outlined a possible mechanism for the learning of such representations based on the fundamental capability to make predictions about the environment and the reactions of others in interactions. Thus, in summary, I have promoted the idea that to make progress towards truly interactive robots, we need to understand how a robot can *learn* to interact, and in what way it can acquire and flexibly apply knowledge about the rules of typical interactions.

References

[1] M. Asada, K. F. MacDorman, H. Ishiguro, and Y. Kuniyoshi. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems*, 37(2–3):185–193, Nov. 2001.

[2] L. W. Barsalou. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(04):577–660, 1999.

[3] E. Bicho, L. Louro, and W. Erlhagen. Integrating verbal and nonverbal communication in a dynamic neural field architecture for Human-Robot interaction. *Frontiers in Neurorobotics*, 4(5), 2010.

[4] A. Fogel. *Change Processes in Relationships: A Relational-Historical Research Approach.* Cambridge University Press, 2006.

[5] V. Gallese. The manifold nature of interpersonal relations: the quest for a common mechanism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431):517–528, Mar. 2003.

[6] N. Hemion. *Building Blocks for Cognitive Robots: Embodied Simulation and Schemata in a Cognitive Architecture.* PhD thesis, Bielefeld University, 2013.

[7] N. J. Hemion, F. Joublin, and K. J. Rohlfing. A competitive mechanism for self-organized learning of sensorimotor mappings. In *Proceedings of the IEEE International Conference on Development and Learning (ICDL)*, Frankfurt am Main, Aug. 2011. IEEE.

[8] A. Howes and R. M. Young. The role of cognitive architecture in modeling the user: Soar's learning mechanism. *Hum.-Comput. Interact.*, 12(4):311–343, Dec. 1997.

[9] J. L. Krichmar and G. M. Edelman. Machine psychology: Autonomous behavior, perceptual categorization and conditioning in a brain-based device. *Cerebral Cortex*, 12(8):818–830, Aug. 2002.

[10] J. E. Laird, A. Newell, and P. S. Rosenbloom. SOAR: an architecture for general intelligence. *Artificial Intelligence*, 33(1):1–64, Sept. 1987.

[11] J. F. Lehman, J. Laird, and P. Rosenbloom. A gentle introduction to soar, an architecture for human cognition. *Invitation to Cognitive Science*, 4, 1996.

[12] M. Minsky. A framework for representing knowledge. In P. H. Winston, editor, *The Psychology of Computer Vision*, pages 211–277. McGraw-Hill, 1974.

[13] A. Morse, J. de Greeff, T. Belpeame, and A. Cangelosi. Epigenetic robotics architecture (ERA). *Autonomous Mental Development, IEEE Transactions on*, 2(4):325–339, Dec. 2010.

[14] J. Piaget. *The origin of intelligence in the child.* Routledge, London; New York, reprint of the 1953 edition, 1997 [1953].

[15] D. E. Rumelhart, P. Smolensky, J. L. McClelland, and G. E. Hinton. Schemata and sequential thought processes in PDP models. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 2: Psychological and Biological Models*, page 7–57. MIT Press, Cambridge, MA, USA, 1986.

[16] R. C. Schank and R. P. Abelson. *Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures.* Artificial Intelligence Series. Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

[17] M. Shanahan. A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition*, 15(2):433–449, 2006.

[18] M. Tomasello. *The Cultural Origins of Human Cognition.* Harvard University Press, 2009.

[19] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. *Science*, 291(5504):599–600, Jan. 2001.

[20] B. Wrede, K. Rohlfing, J. Steil, S. Wrede, P.-Y. Oudeyer, and J. Tani. Towards robots with teleological action and language understanding. In E. Ugur, Y. Nagai, E. Oztop, and M. Asada, editors, *Humanoids 2012 Workshop on Developmental Robotics: Can developmental robotics yield human-like cognitive abilities?*, Osaka, Japon, 2012.