Artificial Curiosity Emerging Human-like Behaviour
A Fundation for Fully Autonomous Cognitive Machines

Dominik Maximilián Ramík, Kurosh Madani and Christophe Sabourin
LISSI Lab. / EA 3956, Sénart-FB Institute of Technology, University Paris-Est Créteil (UPEC),
Campus de Senart, 36-37 rue Georges Charpak, F-77127 Lieusaint, France

Keywords: Artificial Curiosity, Cognitive Machines, Perceptual Curiosity, Epistemic Curiosity, Semantic Knowledge, Learning-by-Interaction, Learning-by-Observation.

Abstract: This paper is devoted to autonomous cognitive machines by mean of the design of an artificial curiosity based cognitive system for autonomous high-level knowledge acquisition from visual information. Playing a chief role as well in visual attention as in interactive high-level knowledge construction, the artificial curiosity (e.g. perceptual and epistemic curiosities) is realized through combining visual saliency detection and Machine-Learning based approaches. Experimental results validating the deployment of the investigated system have been obtained using a humanoid robot acquiring visually knowledge about its surrounding environment interacting with a human tutor. As show the reported results and experiments, the proposed cognitive system allows the machine to discover autonomously the surrounding world in which it may evolve, to learn new knowledge about it and to describe it using human-like natural utterances.

1 INTRODUCTION

Emergence of cognitive phenomena in machines have been and remain active part of research efforts since the rise of Artificial Intelligence (AI) in the middle of the last century, but the fact that human-like machine-cognition is still beyond the reach of contemporary science only proves how difficult the problem is. In fact, if nowadays there are many systems, such as sensors, computers or robotic bodies, that outperform human capacities, nonetheless, none of existing machines or robotic bodies can be called truly intelligent. In other words, machines sharing everyday life with humans are still far away. Somewhat, it is due to the fact that we are still far from fully understanding the human cognitive system. Partly, it is so because if contemporary machines are often fully automatic, they linger rarely fully autonomous in their knowledge acquisition. Nevertheless, the concepts of bio-inspired or human-like machine-cognition remain foremost sources of inspiration for achieving intelligent systems (intelligent machines, intelligent robots, etc…). This is the slant we have taken (e.g. through inspiration from biological and human mechanisms) to investigate the design of a human-like machine-cognition based system able to acquire high-level semantic knowledge from perceptual (namely visual) information. Our main source of inspiration has been the “human’s curiosity” intellectual process for discovering the surrounding world or acquiring new knowledge about it.

It is important to emphasize that the term “cognitive system” means here that characteristics of such a system tend to those of human’s cognitive system. This means that a cognitive system, which is supposed to be able to comprehend the surrounding world on its own, but whose comprehension would be non-human, would afterward be incompetent of communicating about it with its human counterparts. In fact, human-inspired knowledge representation and human-like communication (namely semantic) about the acquired knowledge become key points expected from such a system. To achieve the aforementioned capabilities such a cognitive system should thus be able to develop its own high-level representation of facts from low level visual information (such as image). Accordingly to expected autonomy, the processing from the “sensory level” to the “semantic level” should be performed solely by the robot, without human supervision. However, this does not mean excluding interaction with human, which is, on the contrary, vital for any cognitive system, be it human or

Maximilián Ramík D., Madani K. and Sabourin C..
Artificial Curiosity Emerging Human-like Behaviour - A Fundation for Fully Autonomous Cognitive Machines.
DOI: 10.5220/0004628604070419
In *Proceedings of the 5th International Joint Conference on Computational Intelligence* (NCTA-2013), pages 407-419
ISBN: 978-989-8565-77-8

machine. Thus the investigated system shares its perceptual high-level knowledge of the world with a human tutor by interacting with him. The tutor on his turn shares with the cognitive robot his knowledge about the world by in natural speech (utterances) completing observations made by the robot.

Curiosity is indeed a foremost mechanism among key skills for human cognition. It may play the role of an appealing source for conceiving artificial systems that gather knowledge autonomously. We will devote a discussion on this purpose further in this paper. Nevertheless, we have taken into consideration this enticing cognitive skill making it our principle foundation in investigated concept. The present paper is devoted to the description of a cognitive system based on artificial curiosity for high-level knowledge acquisition from visual information. The goal of the investigated system is to allow the machine (such as a humanoid robot) to anchor the heard terms to its visual information and to flexibly shape this association according to its budding knowledge about the observed items within its surrounding world. In other words, the presented system allows the machine to observe, to learn and to interpret the world in which it evolves, using appropriate terms from human language, while not making use of a priori knowledge. This is done by word-meaning anchoring based on learning by observation stimulated (steered) by artificial curiosity and by interaction with the human tutor. Our model is closely inspired by juvenile learning behaviour of human infants (Yu, 2005), (Waxman,2009). By analogy with natural curiosity the artificial curiosity has been founded on two cognitive levels. The first ahead of reflexive visual attention plays the role of perceptual curiosity and the second coping with intentional learning-by-interaction undertakes the role of epistemic curiosity.

The present paper is further organized as follow. Next section is dedicated to a brief overview of existing techniques in autonomous learning and knowledge acquisition, especially in robotics systems. Section three elucidates theoretical aspects of the investigated cognitive system. Section four briefly runs through perceptual curiosity, relying on our previously published works on salient vision. Section five details the higher-level cognitive layer and provides its validation. Section six provides details about deployment of the system on a humanoid robot in real world conditions. Finally a conclusion and a perspective on future directions close this paper.

2 BRIEF OVERVIEW OF RELATED WORKS

Before running through a brief synopsis of already accomplished works and available techniques relating the purpose of this paper, it is pertinent to note that the cognition and related aspect cover an extremely extensive spectrum of competencies and regroup a huge hoard of multi-disciplinary works. Thus, it is neither the purpose of this section nor our intent to fully overview the colossal amount of research works linking different parts of the presented work. That is why, in this section we will focus on research efforts that have played, in some way, an influential role for achieving the presented work or on those closely related to its subject. We therefore focus on cognitive systems, perceptual curiosity (notably visual saliency) and on works concerning knowledge acquisition.

In the present work the term “cognition” is considered as human-like knowledge based functionality of machines. A machine (or a robot) responding correctly to such challenge cannot rely only on a priori knowledge that has been stored in it, but should be able to learn on-line from environment where it evolves by interaction with the people it encounters in that environment. On this subject, the reader may refer to (Kuhn et al., 1995), a monograph on knowledge acquisition strategies and to (Goodrich and Schultz, 2007) giving a survey on human-robot interaction and learning and to (Coradeschi and Saffiotti, 2003) providing an overview of the anchoring problem. In (Madani and Sabourin, 2011), a multi-level cognitive machine-learning based concept for human-like “artificial” walking is proposed. Authors define two kinds of cognitive functions: the “unconscious cognitive functions” (UCF), identified as “instinctive” cognition level handling reflexive abilities, and “conscious cognitive functions” (CCF), distinguished as “intentional” cognition level handling thought-out abilities. In (Madani, 2012) authors focus the concept of Artificial Awareness based on visual saliency with application to humanoid robot's awareness.

The autonomous learning benefiting from interaction with humans will inherently require the machine’s ability of learning without explicit “negative training set” (or negative evidence) and from a relatively small number of samples. This important capacity is observed in children learning the language and is discussed in (Bowerman, 1983). The problem of autonomous learning has been addressed on different degrees in several works. For example, in (Greeff et al., 2009) a computational model of word-meaning acquisition by interaction is presented. In (Saunders, 2010) a humanoid robot is taught by a human tutor to associate simple shapes to human lexicon in an interactive way. A more advanced work on autonomous robot learning using a weak form of interaction with the tutor has been recently presented in (Araki et al., 2011). Another interesting approach to autonomous learning of visual concepts in robots has been published in (Skocaj et al., 2011). Authors show capacity of their robotic platform to engage in different kinds of learning in interaction with a human tutor.

Concerning the use of curiosity in machine-cognition, by observing the state of the art it may be concluded that the curiosity is usually used as an auxiliary, single-purpose mechanism, instead of being the fundamental basis of the knowledge acquisition. In (Ogino et al., 2006), a lexical acquisition model is presented combining more traditional approaches with the concept of curiosity to alternate the attention of the learning robot. To our best knowledge there is no work to date which considers curiosity in context of machine cognition as a drive for knowledge acquisition on both low (perceptual) level and high ("semantic") level of the system, as it is described in this chapter.

Visual saliency (also referred in literature as visual attention, unpredictability or surprise) is described as a perceptual quality that makes a region of image stand out relative to its surroundings and to capture attention of observer (Achanta et al., 2009). The inspiration for the concept of visual saliency comes from the functioning of early processing stages of human vision system and is roughly based on previous clinical research. In early stages of the visual stimulus processing, human vision system first focuses in an unconscious, bottom-up manner, on visually attractive regions of the perceived image. The visual attractiveness may encompass features like intensity, contrast and motion. Although there exist solely biologically based approaches to visual saliency computation, most of the existing works do not claim to be biologically plausible. Instead, they use purely computational techniques to achieve the goal. One of the first works using visual saliency in image processing has been published by (Itti et al., 1998). Authors use a biologically plausible approach based on a centre-surround contrast calculation using Difference of Gaussians. Published more recently, other common techniques of visual saliency calculation include graph-based random walk (Harel et al., 2007), centre-surround feature distances (Achanta et al., 2008), multi-scale contrast, centre-surround histogram and color spatial distribution or features of color and luminance (Liu, 2008). A less common approach is described in (Liang et al., 2012). It uses content-sensitive hyper-graph representation and partitioning instead of using more traditional fixed features and parameters for all images.

3 CURIOSITY BASED ARTIFICIAL INTELLECT

3.1 Concept and Role of Artificial Curiosity

As it has already been mentioned, "curiosity" is a key skill in human cognitive ability for acquiring knowledge. Thus it is an appealing concept in conceiving artificial systems supposed to gather knowledge autonomously. So, before exposing the investigated system let us focus on curiosity in more depth.

Berlyne (Berlyne, 1954) addresses the concept of human's curiosity by splitting up the curiosity into two kinds. The first, so-called "perceptual curiosity", leads to increased perception of stimuli. It is a lower level function, relating perception of new, surprising or unusual sensory inputs. It contrasts to repetitive or monotonous perceptual experience. The other one, so called "epistemic curiosity", is related to the desire for knowledge that motivates individuals to learn new ideas, to eliminate information-gaps, and to solve intellectual problems (Litman, 2008).

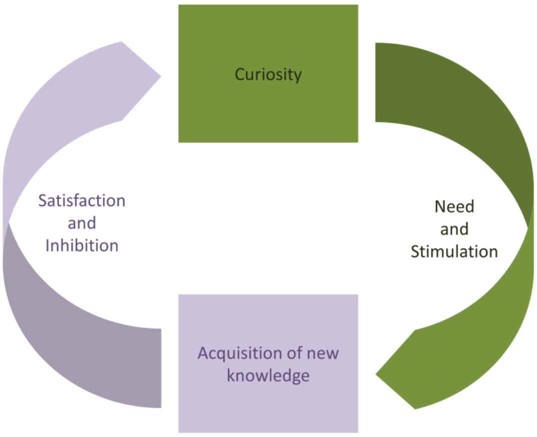


Figure 1: General concept of knowledge acquisition's regulation by curiosity in human cognition.

It also seems that it acts to stimulate long-term memory in remembering new or surprising information (Kang et al., 2009).Without striving for biological plausibility, the above-mentioned gives an important motivation for building of our system: curiosity stimulates acquisition of new knowledge and in turn the newly learned knowledge whips up or appeases the curiosity. In other words, it is the curiosity, which motivates and regulates any action of the system. Figure 1 depicts the above-formulated concept.

3.1.1 Perceptual Curiosity and Visual Saliency

In their perception, humans rely strikingly much on vision. It is then pertinent to consider chiefly the visual information and its learning processes. Thus, it appears appropriate here to draw inspiration from studies on human infants learning by demonstration. Experiments in (Brand et al., 2002) show that it is the explicitness or exaggeration of an action that helps a child to understand, what is important in the actual context of learning. It may be generalized, that it is the saliency (in terms of motion, colors, etc.) that lets the pertinent information "stand-out" from the context and become "surprising" (Wolfe, 2004). We argue that in this context the visual saliency may be helpful to enable unsupervised extraction and subsequent learning of a previously unknown object by a machine. In other words, perceptual curiosity has been realized through a saliency detection approach.

3.1.2 Epistemic Curiosity and Learning-by-Observation and Interaction

Epistemic curiosity stimulates the high level knowledge acquisition mechanism constructing new semantic knowledge and to fill the gaps of missing knowledge. Thus, epistemic curiosity operates inherently at "conscious" cognitive level, as it requires an intentional search and premeditated interaction with the environment. This mechanism allows the machine to learn abstract (e.g. insubstantial) knowledge, after interpreting the world in which it evolves, by using appropriate terms from human language. It is important to stress that this is done without making use of a priori knowledge. The task is realized by word-meaning anchoring based on learning-by-observation and by interaction with its human counterpart. The model is closely inspired by learning process of human infants.

The machine shares its perception of the surrounding world with the human (tutor) and interacts with him. The tutor on his turn shares with the machine his knowledge about the world within the form of natural speech (utterances) accompanying machine's observations and completing its knowledge about the perceived reality. In other words, of such a high-level cognitive mechanism is to allow the machine to anchor the heard terms to its sensory-motor experience and to flexibly shape this anchoring accordingly to its growing knowledge about the world. The described mechanism can play a key role in linking object extraction and learning techniques on one side, and ontologies on the other side. The former ones are closely related to perceptual reality, but are unaware of the meaning of objects they identify. While the latter ones are able to represent complex semantic knowledge about the world, but, they are unaware of the perceptual reality of concepts they are handling.

3.2 Architecture of the Curiosity based Artificial Cognitive Intellect

As depicted in figure 2, the general architecture of the investigated artificial intellectual system is organized around four main modules, derived from needs outlined previously and from what has been previously mentioned about the role of curiosity. The "Behaviour Control" unit shapes the overall coherent (intelligent) behaviour of the machine by collecting results from the three other units and by providing information and requests issued from surrounding environment. The "Navigation" unit is in charge of machines apposite evolution in its surrounding world. The task of the "Communication" unit is to allow communicating this knowledge to the outer world and to handle inputs from humans and transfer them into a machine readable form. It enables the system to communicate in two ways with other actors, be it similar intelligent machines or human beings. Finally, the task of "Knowledge Acquisition" unit is knowledge gathering and handling, derivation of high-level representation from low-level sensory data and construction of semantic relationships from interpretation of perceived information.

Accordingly to (Madani and Sabourin, 2011) and as in (Madani et al., 2012), the general concept of "Knowledge Acquisition" unit could be depicted as shown in figure 3. It includes one unconscious visual level containing a number of UCF and one conscious visual level which may contains a number of CCF. The knowledge extraction from visual pattern follows the process involving both kinds of aforementioned curiosity (“perceptual curiosity” and the “epistemic curiosity”.

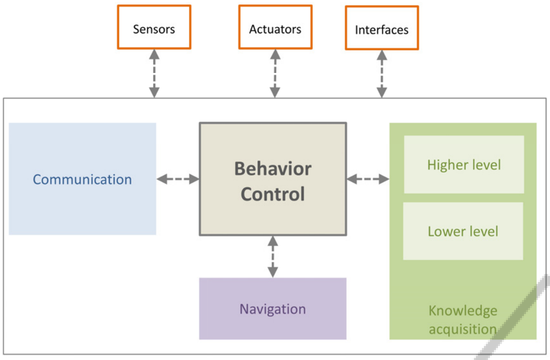


Figure 2: Block diagram showing the general architecture of the investigated artificial intellectual system.

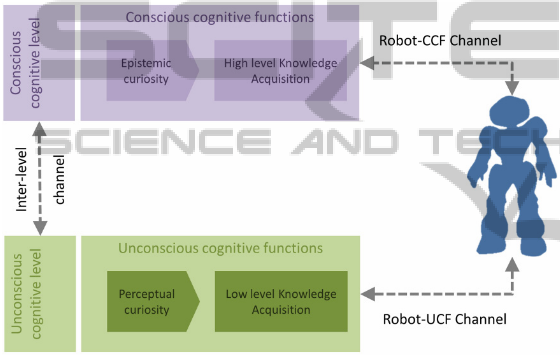

Conscious cognitive level
Conscious cognitive functions
Epistemic curiosity
High level Knowledge Acquisition
Robot-CCF Channel
Inter-level channel
Unconscious cognitive level
Unconscious cognitive functions
Perceptual curiosity
Low level Knowledge Acquisition
Robot-UCF Channel

Figure 3: Block diagram of “Knowledge Acquisition” unit and places of the perceptual and the epistemic curiosities.

The perceptual curiosity motivates or stimulates what we call the low level knowledge acquisition and concerns “reflexive” (unconscious) processing level. It seeks “surprising” or “attention-drawing” information in given visual data. The task of the perceptual curiosity is realized by perceptual saliency detection mechanisms. This gives the basis for operation of high-level knowledge acquisition, which is stimulated by epistemic curiosity. Being previously defined as the process, that motivate to “learn new ideas and solving intellectual problems”, the epistemic curiosity is here the motor of: learning new concepts based on what has been gathered on the lower-level and eliminating information gaps by encouraging an active search for the missing information.

4 PERCEPTUAL CURIOSITY THROUGH VISUAL SALIENCY

As mentioned in previous section, the perceptual curiosity relates visual attention and thus could be realized through the saliency detection approach. However, the exiting salient objects’ detection approaches as well as those connecting the detection and recognition techniques used by those approaches rely on human made databases, requiring a substantial time and a skilled human expert. Thus, a fully autonomous machine vision system, aiming recognizing salient objects on its own, could not be achieved with the above-mentioned techniques. Motivated by the mentioned shortcoming regarding existing object recognition methods, we have proposed earlier an intelligent Machine-Vision system able to detect and to learn autonomously individual objects within real environment. The approach has been detailed in (Ramik, 2011-a) and (Ramik, 2011-b) using an architecture following “cognitive” frame described in (Madani, 2011) and (Madani, 2012). Its key capacities are: autonomous extraction of multiple objects from raw unlabeled camera images, learning of those objects autonomously and recognition of the learned objects in different conditions or in different visual contexts.

Allowing the machine to learn and to recognize

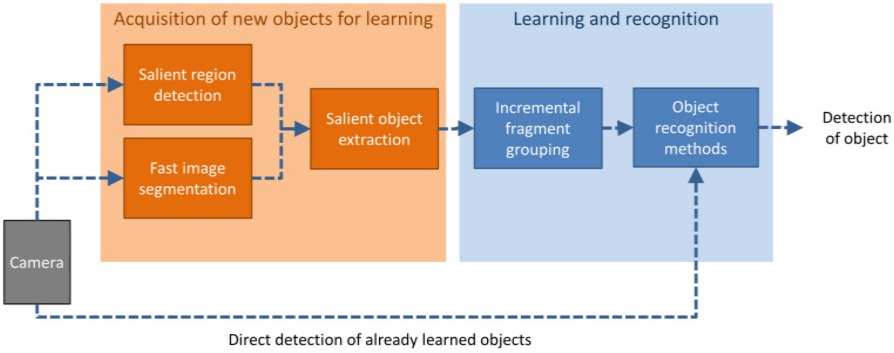
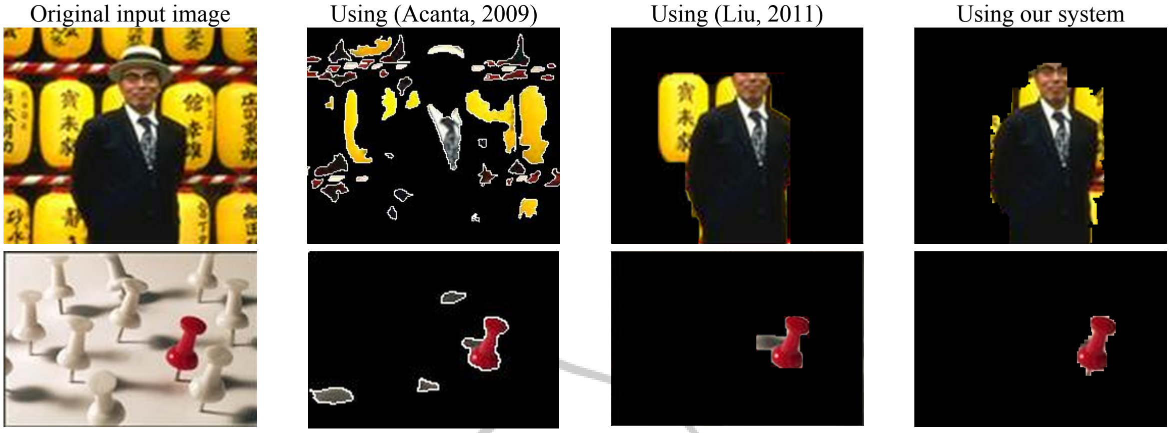

Acquisition of new objects for learning
Salient region detection
Fast image segmentation
Salient object extraction
Learning and recognition
Incremental fragment grouping
Object recognition methods
Detection of object
Camera
Direct detection of already learned objects

Figure 4: Block diagram of visual saliency system.



Original input image

Using (Acanta, 2009)

Using (Liu, 2011)

Using our system

Figure 5: Comparison of different salient object detection algorithms. 1st column: original image, 2nd column: results using (Achanta et al., 2009), 3rd column: results using (Liu et al., 2011) and 4th column: results of our approach.

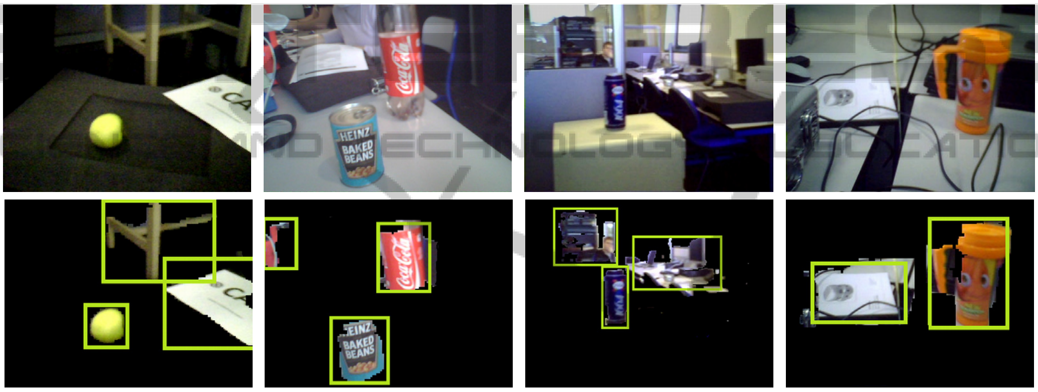


Figure 6: Samples of real environment issued images (upper row) and salient objects detected by our algorithm, marked by green rectangles (lower row).

Allowing the machine to learn and to recognize objects encountered in its surrounding environment in a completely automated manner, the designed object learning system consists of several units which collaborate together. Figure 4 depicts the block-diagram of the system showing different units and their relations. Two main parts may be identified, each one containing several processing units. The first part, labelled “Acquisition of new objects for learning” takes a raw image from the camera, detects visually important objects on it and extracts them so that they can be used as prospective samples for learning. In parallel the input image is segmented and split into a set of segments according to the chromatic surface properties.

The algorithm is shown to be robust to common illumination effects like shadows and reflections, which helps our system to cope with real illumination conditions. Finally, combining results of the two aforementioned units, the “Salient object extraction” unit extracts segments found on salient regions and constructs the salient objects from the input image.

The second part incrementally clusters the detected salient objects, learns new salient objects and handles the recognition of previously learned objects (in recognition phase). In fact, the extracted salient objects are fed into the “Incremental fragment grouping” unit. Here, an on-line classification is performed on each object by a set of weak classifiers and incrementally groups containing the same object extracted from different images are formed. These groups can be then used as a kind of visual memory of visual database describing each of the extracted objects. This alone would be enough for recognition of each already seen object, if it was ensured that each particular object will be found in the same visual context next time it is encountered by our system. However, such hypothesis (e.g. expectation) is clearly too restrictive for a system that aims to recognize the once learned objects in any conditions or contexts. That is why the last unit of the system, tagged "Object recognition methods", is added. Its role is, by employing existing object recognition algorithms, to learn from the visual database built by "incremental fragment grouping" unit and to recognize those objects regardless to their saliency in new settings. Thus the already learned objects can be recognized directly from the input image. Figures 5 and 6 give examples of salient objects' detection obtained by the presented system compared to two already existing techniques (figure 5) as well as examples obtained from observing the real environment (figure 6).

With respect to the expected goal requiring real-time processing skills, the system is designed with emphasis on on-line and real-time operation. Moreover, the system itself is however not limited to mobile platforms and can be amply used in context of various other applications (as those dealing with sensor networks, intelligent houses etc...).

5 HIGH-LEVEL KNOWLEDGE ACQUISITION

The problem of learning brings an inherent problem of distinguishing the pertinent sensory information (the one to which the tutor is referring) and the impertinent one. It indeed is a paradox, but in contrary to what one may believe, sensors provide generally too much data input: a lot more than the amount effectively needed. It is the task of higher structures (e.g. an attention system or in general a machine learning system adapted to this task) to draw the attention to particular features of the data, which are pertinent in context of a particular task. The solution to this task is not obvious even if we achieve joint attention in the robot. This is illustrated on figure. 7. Let us consider a robot (machine) learning a single type of features, e.g. for example colors. If a tutor points to one object (e.g. for example a red flower) among many others, and describes it by saying: "The flower is red!", the robot still has to distinguish which of the several colors and shades, found on the concerned object, the tutor is referring to. This step is an inevitable one before beginning the learning itself. In traditional learning systems, such task-relevant (i.e. pertinent) information is extracted by a human expert. In a system capable of autonomous learning, however, this has to be done in an automated way and without recourse to human-extracted features.

Figure 8 gives the bloc-diagram of key operations flow of the system proposed in this section. As it could be seen in figure 7, sensor data bring inherently both pertinent and impertinent information mixed up. To achieve correct detection of pertinent information in spite of such an uncertainty, we adopt the following strategy. The system extracts features from important objects found in the scene along with words the tutor used to describe the presented objects. Then, the system generates its beliefs about which word could describe which feature. The beliefs are seen as organisms in a genetic algorithm. Here, the appropriate fitness function is of major importance. To calculate the fitness, a classifier is trained based on each belief about the world. Using it, the cognitive system undertakes to interpret the objects it has already seen. The utterances pronounced by the human tutor in presence of each already seen object are compared with the machine's utterances used to describe it based on its current belief. The closer the machine's description is to the one given by the human, the higher the fitness is.

5.1 Observation and Interpretation

The present sub-section explains how observed (e.g. visual) information is interpreted by the presented system. For, this, let us suppose that visual information is acquired through appropriate sensor (for example a camera, etc…), which makes the system able to observe the surrounding world. This means that the observed world is represented as a set of features $I = \{i_1, i_2, \cdots, i_k\}$. Let us also suppose that each time the machine makes an observation O, a human tutor gives it a set of utterances U_H describing important objects found currently in the observed world. Let us denote U the set of all utterances ever given about the world. Accordingly to what has been introduced at the beginning of this section, the goal for the machine is to discriminate the pertinent information from the impertinent one and to correctly map the utterances to appropriate perceived features. In other words, the machine has to establish a word-meaning relationship between the uttered words and its own perception about the observed information. The machine is further allowed to interact with the human in order to clarify and verify its interpretations, following the stimulation of its epistemic curiosity.

For this purpose, we define an observation O as an ordered pair $o = \{I_l, U_H\}$, where $I_l \subseteq I$ stands for the set of features obtained from observation and $U_H \subseteq U$ is the set of utterances given in the

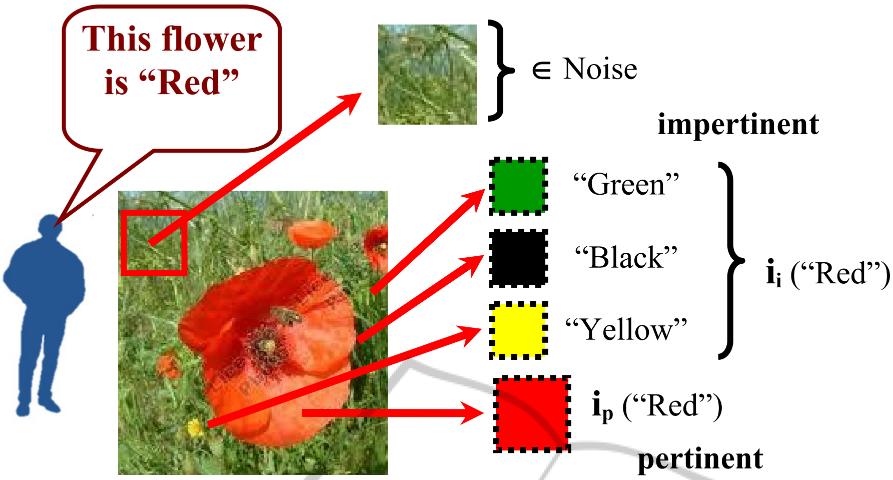


Figure 7: A human would describe this flower as being "red" in spite of the fact, that this is not its only color.

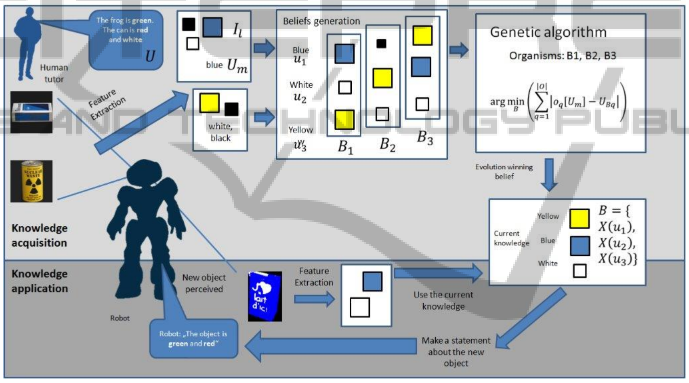


Figure 8: The proposed system' operational flow showing learning process of single-type of features. The example shows the context of a particular learning task, i.e. color learning and interpretation.

context of that observation. Following Eq. (1) I_l is a sum of all the pertinent information i_p for a given utterance $u \in U$ (i.e. features that can be described as u in the language used for communication between the human and the robot), all the impertinent information i_i (i.e. features that are not described by the given u, but might be described by another $u_i \in U$) and noise ε.

$$I_l = \bigcup_{U_H} i_p(u) + \bigcup_{U_H} i_i(u) + \varepsilon \qquad (1)$$

Let us define in a general way, an interpretation $X(u)$ of an utterance u as an ordered pair $X(u) = \{u, I_j\}$ (where $I_j \subseteq I$), which denotes that a sub-set of features I_j of I is interpreted as u. Then a belief is defined accordingly to Eq. (2) as an ordered set of $X(u)$ interpreting utterances u, where $n = |U|$. Belief B is a mapping (relation) from the set of U to I. All members of U map to one or more members of I and no two members of U map to the same member of I.

$$B = \{X(u_1), \cdots, X(u_n)\} \qquad (2)$$

According to Eq. (3) one can determine the belief B, which interprets in the most coherent way the observations made so far. It is done by looking for such a belief, which minimizes across all the observations $o_q \in O$ the difference between the utterances $U_q \subset U_H$ made by human on each particular observation $o_q \in O$, and those composed the machine (denoted U_{Rq}), by using the belief B on the same observation. In other words, we are looking for a belief B, which would make the machine describe a particular sight with utterances as close as possible to those that would make a human on the same prospect.

$$\arg\min_{B} \left(\sum_{q=1}^{|O|} \left| U_{Hq} - U_{Rq} \right| \right) \qquad (3)$$

5.2 Evolutionary Searching for Most Coherent Interpretation

The system has to look for a belief B, which would make the robot describing a particular scene with utterances as close and as coherent as possible to those made by a human on the same scene. For this purpose, instead performing the exhaustive search over all possible beliefs, we propose to search for a suboptimal belief by means of a genetic algorithm. For doing that, we assume that each organism within it has its genome constituted by a belief, which, results into genomes of equal size $|U|$ containing interpretations $X(u)$ of all utterances from U. The task of coherent belief generation is to generate beliefs, which are coherent with the observed reality. In our genetic algorithm, the genomes' generation is a belief generation process generating genomes (e.g. beliefs) as follows. For each interpretation $X(u)$ the process explores whole the set O. For each observation $o_q \in O$, if $u \in U_{Hq}$ then features $i_q \in I_q$ (with $I_q \subseteq I$) are extracted. As described in (1), the extracted set of features contains as well pertinent as impertinent features.

The coherent belief generation is done by deciding, which features $i_q \in I_q$ may possibly be the pertinent ones. The decision is driven by two principles. The first one is the principle of "proximity", stating that any feature i is more likely to be selected as pertinent in the context of u, if its distance to other already selected features is comparatively small. The second principle is the "coherence" with all the observations in O. This means, that any observation $o_q \in O$, corresponding to $u \in U_{Hq}$, has to have at least one feature assigned into I_q of the current $X(u) = \{u, I_q\}$.

To evaluate a given organism, a classifier is trained, whose classes are the utterances from U and the training data for each class $u \in U$ are those corresponding to $X(u) = \{u, I_q\}$, i.e. the features associated with the given u in the genome. This classifier is used through whole set O of observations, classifying utterances $u \in U$ describing each $o_q \in O$ accordingly to its extracted features. Such a classification results in the set of utterances U_{Rq} (meaning that a belief B is tested regarding the qth observation). Block-diagram of described genetic algorithm's workflow is given by figure 9.

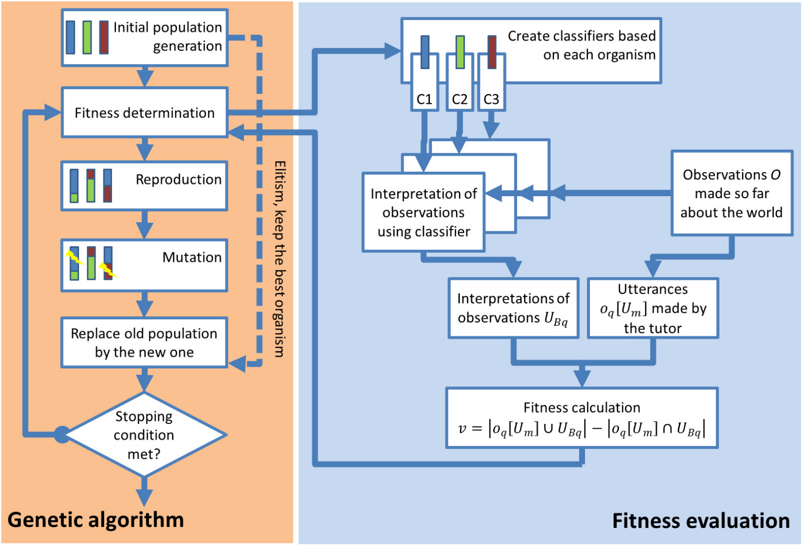

Initial population generation
Fitness determination
Reproduction
Mutation
Replace old population by the new one
Stopping condition met?
Elitism, keep the best organism
Genetic algorithm
Create classifiers based on each organism
C1
C2
C3
Interpretation of observations using classifier
Observations O made so far about the world
Interpretations of observations U_{Bq}
Utterances $o_q[U_m]$ made by the tutor
Fitness calculation
$v = |o_q[U_m] \cup U_{Bq}| - |o_q[U_m] \cap U_{Bq}|$
Fitness evaluation

Figure 9: Bloc diagram of described genetic algorithm's workflow. The left part describes the genetic algorithm itself, while the right part focuses on the fitness evaluation workflow.

$$D(\nu)=\frac{1}{1+\nu} \tag{4}$$

$$\nu=\left|U_{Hq}\bigcup U_{Rq}\right|-\left|U_{Hq}\bigcap U_{Rq}\right| \tag{5}$$

5.3 Role of Human-machine Interaction

In our approach, the learning by interaction is carried out in two kinds of interactions: human-to-machine and machine-to-human. The human-to-machine interaction is activated anytime the machine interprets wrongly a given word world. When the human receives a wrong response, he provides the machine a new observation by uttering the desired interpretation. Then system searches for a new interpretation of the world conformably to this new observation. The machine-to-human interaction may be activated when the robot attempts to interpret a particular feature classified with a very low confidence. Led by the epistemic curiosity, the machine asks its human counterpart to make an utterance about the uncertain observation. If machine's interpretation is not conforming to the utterance given by the human, this observation is recorded as a new knowledge and a search for the new interpretation is started.

6 VALIDATION AND EXPERIMENTAL RESULTS

The validation of the proposed system has been performed on the basis of both simulation of the designed system as by an implementation on a real humanoid robot. A video capturing different parts of the experiment may be found online on: http://youtu.be/W5FD6 zXihOo. As real robot we have considered NAO robot (a small humanoid robot from Aldebaran Robotics) which provides a number of facilities such as onboard camera (vision), communication devices and onboard speech generator. The fact that the above-mentioned facilities been already available offers a huge save of time, even if those faculties remain quite basic in that kind of robots.

Although the usage of the presented system is not specifically bound to humanoid robots, it is pertinent to state two main reasons why a humanoid robot is used for the system's validation. The first reason for this is that from the definition of the term "humanoid", a humanoid robot is aspired to make its perception close to the human's one, entailing a more human-like experience of the world. This is an important aspect to be considered in context of sharing knowledge between a human and a robot. The second reason is that humanoid robots are specifically designed to interact with humans in a "natural" way by using e.g. a loudspeaker and microphone set in order to allow for a bi-directional communication with human by speech synthesis and speech analysis and recognition. This is of importance when speaking about a natural human-robot interaction during learning.

6.1 Simulation based Validation

The simulation based validation finds its pertinence in assessment of the investigated cognitive-system's performances. In fact, due to difficulties inherent to organization of strictly same experimental protocols on different real robots and within various realistic contexts, the simulated validation becomes an appealing way to ensure that the protocol remains the same. For simulation based evaluation of the behaviour of the above-described system, we have considered color names learning problem. In everyday dialogs, people tend to describe objects, which they see, with only a few color terms (usually only one or two), although the objects in itself contains many more colors. Also different people can have slightly different preferences on what names to use for which color. Due to this, learning color names is a difficult task and it is a relevant sample problem to test our system.

In the simulated environment, images of real-world objects were presented to the system alongside with textual tags describing colors present on each object. The images were taken from the Columbia Object Image Library (COIL) contains 1000 color images of different views of 100 objects database. Five fluent English speakers were asked to

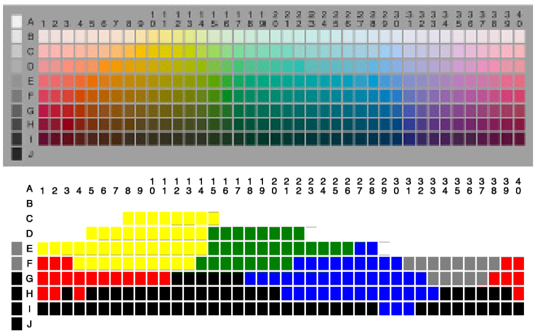


Figure 10: Original WCS table (upper image), its system's made interpretation (lower image).

describe each object in terms of colors. We restricted the choice of colors to “Black”, “Gray”, “White”, “Red”, “Green”, “Blue” and “Yellow”, based on the color opponent process theory (Schindler, 1964).

The tagging of the entire set of images was highly coherent across the subjects. In each run of the experiment, we have randomly chosen a tagged set. The utterances were given in the form of text extracted from the descriptions. The object was accepted as correctly interpreted if the system’s and the human’s interpretations were equal. The rate of correctly described objects from the test set was approximately 91%. Figure 10 gives the result of interpretation by the system of the colors of the WCS table.

6.2 Real Robot based Validation

The designed system has been implemented on NAO robot (from Aldebaran Robotics). It is a small humanoid robot which provides a number of facilities such as onboard camera (vision), communication devices and onboard speech generator. The fact that the above-mentioned facilities been already available offers a huge save of time, even if those faculties remain quite basic in that kind of robots. If NAO robot integrates an onboard speech-recognition algorithm (e.g. some kind of speech-to-text converter) which is sufficient for “hearing” the tutor, however its onboard speech generator is a basic text-to-speech converter. It is not sufficient to allow the tutor addressing the robot in natural speech. To overcome NAO’s limitations relating this purpose, the TreeTagger tool was used in combination with robot's speech-recognition system to obtain the part-of-speech information from situated dialogs. Developed by the ICL at University of Stuttgart, available online at: http://www.ims.uni-stuttgart.de/projekte/corplex/ TreeTagger. Standard English grammar rules were used to determine whether the sentence is demonstrative, descriptive or an order. To communicate with the tutor, the robot used its text-to-speech engine.

A number of every-day objects have been collected for purposes of the experiment. They have been randomly divided into two sets for training and for testing. The learning set objects were placed around the robot and then a human tutor pointed to each of them calling it by its name. Using its 640x480 monocular color camera, the robot discovered and learned the objects around it by the salient object detection approach we have described earlier. Here, this approach has been extended by detecting the movement of the tutor's hand to achieve joint attention. In this way, the robot was able to determine what object the tutor is referring to and to learn its name. Figure 11 shows two photographs of the above-reported experimental validation. Figure 12 shows examples of two extracted objects and the robot’s interpretation of the concerned objects.

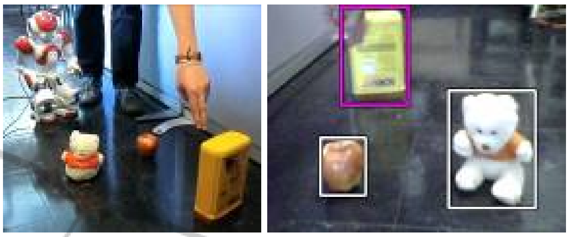


Figure 11: Experimental setup showing the tutor pointing a yellow chocolate box which has been seen, interpreted and learned (by the robot) in term of colors then asking the robot to describe the chosen object (lower left-side) and the ground truth detected objects as the robot perceives them (lover right-side).

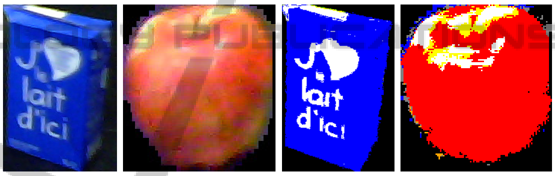


Figure 12: Two objects extracted from robot's surroundings. Right: the original image, left: features interpreted. For the “apple”, the robot's given description was “the object is red”. For the box, the description was “the object is blue and white”.

During the experiment, the robot has been asked to learn a subset among the 25 considered objects: in term of associating the name of each detected object to that object. At the same time, a second learning has been performed involving the interaction with the tutor who has successively pointed the above-learned objects describing (e.g. telling) to the robot the color of each object. Here-bellow an example of the Human-Robot interactive learning is reported:

- **Human** [*pointing a red aid-kit*]: “This is a first-aid-kit!”
- **Robot**: “I will remember that this is a first-aid-kit.”
- **Human**: “It is red and white”.
- **Robot**: “OK, the first-aid-kit is red and the white.”

After learning the names and colors of the discovered objects, the robot is asked to describe a number of objects including as well some of already learned objects but in different posture (for example the yellow box presented in reverse posture) as a number of still unseen objects (as for example a red apple or a white teddy-bear). The robot has successfully described, in a coherent linguistics, the presented seen and unseen objects. Here-bellow is the Human-Robot interaction during the experiment:

- **Human** [*pointing the unseen white teddy-bear*]: “Describe this!”
- **Robot**:: “It is white!”
- **Human** [*pointing the already seen, but reversed, yellow box*]: “Describe this!”
- **Robot**: “It is yellow!”
- **Human** [*pointing the unseen apple*]: “Describe this!”
- **Robot**: “It is red!”

7 CONCLUSIONS

This paper has presented, discussed and validated a cognitive system for high-level knowledge acquisition based on the notion of artificial curiosity. Driving as well the lower as the higher levels of the presented cognitive system, the emergent artificial curiosity allow such a system to learn in an autonomous manner new knowledge about unknown surrounding world and to complete (enrich or correct) its knowledge by interacting with a human. Experimental results, performed as well on a simulation platform as using the NAO robot show the pertinence of the investigated concepts as well as the effectiveness of the designed system. Although it is difficult to make a precise comparison due to different experimental protocols, the results we obtained show that our system is able to learn faster and from significantly fewer examples, than the most of more-or-less similar implementations.

Based on obtained results, it is thus justified to say, that a robot endowed with such artificial curiosity based intelligence will necessarily include autonomous cognitive capabilities. With respect to this, the further perspectives will focus integration of the investigated concepts in other kinds of machines, such as mobile robots. There, it will play the role of an underlying system for machine cognition and knowledge acquisition.

REFERENCES

Achanta, Estrada, F., Wils, P., Susstrunk, S., 2008. Salient Region Detection and Segmentation, Proc. of International Conference on Computer Vision Systems (ICVS '08), vol. 5008, LNCS, Springer Berlin / Heidelberg, 66_75.

Achanta, R., Hemami, S., Estrada, F., Susstrunk, S., 2009 Frequency-tuned Salient Region Detection, Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR).

Araki, T., Nakamura, T., Nagai, T., Funakoshi, K., Nakano, M. and Iwahashi, N., 2011 'Autonomous acquisition of multimodal information for online object concept formation by robots', IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, 1540-1547.

Berlyne, D. E., 1954, 'A theory of human curiosity', *British Journal of Psychology*, vol. 45, no. 3, August, pp. 180-191.

Bowerman, M. , 1983, 'How Do Children Avoid Constructing an Overly General Grammar in the Absence of Feedback about What is Not a Sentence?', *Papers and Reports on Child Language Development*.

Brand, R. J., Baldwin, D. A. and Ashburn, L. A., 2002 'Evidence for 'motionese': modifications in mothers infant-directed action', Developmental Science, 72-83.

Coradeschi, S. and Saffiotti, A., 2003, 'An introduction to the anchoring problem', *Robotics and Autonomous Systems*, vol. 43, pp. 85-96.

Goodrich, M. A. and Schultz, A. C., 2007, 'Human-robot interaction: a survey', *Foundations and trends in human computer interaction*, vol. 1, jan, pp. 203-275.

Greeff, J. D., Delaunay, F. and Belpaeme, T., 2009, 'Human-Robot Interaction in Concept Acquisition: a computational model', Proceedings of the 2009 IEEE *8th International Conference on Development and Learning, Washington*, 1-6.

Harel, J., Koch, C., Perona, P., 2007, Graph-based visual saliency, *Advances in Neural Information Processing Systems* 19, 545_552.

Itti, L., Koch, C. , Niebur, E., 1998, A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 1254_1259.

Kang, M. J. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T. T. and Camerer, C. F., 2009, 'The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory', *Psychological science*, vol. 20, no. 8, August, pp. 963-973, Available: 1467-9280.

Kuhn, D., Garcia-Mila, M., Zohar, A. and Andersen, C., 1995, 'Strategies of knowledge acquisition', *Society for Research in Child Development Monographs*, vol. 60 (4), no. 245.

Liang, Z., Chi, Z., Fu, H., Feng, D., 2012, Salient object detection using content-sensitive hypergraph representation and partitioning, Pattern Recogn. 45 (11), 3886_3901.

Litman, J. A., 2008, 'Interest and deprivation factors of epistemic curiosity', *Personality and Individual Differences*, vol. 44, no. 7, pp. 1585-1595.

Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N. , Tang X., Shum, H.-Y., 2011, Learning to Detect a Salient Object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2), 353_367.

Madani, K. and Sabourin, C., 2011, 'Multi-level cognitive machine-learning based concept for human-like "artificial" walking: *Application to autonomous stroll of humanoid robots. Neurocomputing*, Vol. 74, 1213-1228.

Madani, K., Ramik, D. M. and Sabourin, C., 2012, Multi-level cognitive machine-learning based concept for

Artificial Awareness: application to humanoid robot's awareness using visual saliency”, J. of Applied *Computational Intelligence and Soft Computing,. DOI*: 10.1155/2012/354785.

Ogino, M., Kikuchi, M. and Asada, M., 2006, 'How can humanoid acquire lexicon? active approach by attention and learning biases based on curiosity', IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, 3480-3485.

Ramík, D. M., Sabourin, C. and Madani K., 2011-a, 'A Real-time Robot Vision Approach Combining Visual Saliency and Unsupervised Learning', *Proc. of 14th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, Paris, 241-248.

Ramík, D. M., Sabourin, C. and Madani, K., 2011-b, 'Hybrid Salient Object Extraction Approach with Automatic Estimation of Visual Attention Scale', Proc. of *7th International Conference on Signal Image Technology & Internet-Based Systems*, Dijon, 438-445.

Saunders, J., Nehaniv, C. L. and Lyon, C., 2010, 'Robot learning of lexical semantics from sensorimotor interaction and the unrestricted speech of human tutors', *Second International Symposium on New Frontiers in Human-Robot Interaction*, Leicester, 95-102.

Schindler, M. and von Goethe, J. W., 1964, Goethe’s theory of colour applied by Maria Schindler. New Knowledge Books, East Grinstead, Eng.

Skocaj, D., Kristan, M., Vrecko, A., Mahnic, M., Janicek, M., Kruijff, G.-J.M., Hanheide, M., Hawes, N., Keller, T., Zillich, M. and Zhou, K., 2011, 'A system for interactive learning in dialogue with a tutor', IEEE/RSJ *International Conference on Intelligent Robots and Systems* IROS 2011, San Francisco, 3387-3394.

Waxman, S. R. and Gelman, S. A., 2009, 'Early word-learning entails reference, not merely associations', *Trends in cognitive science*, vol. 13, jun, pp. 258-263.

Wolfe, J. M. and Horowitz, T. S., 2004, 'What attributes guide the deployment of visual attention and how do they do it?', Nature Reviews Neuroscience, 495-501.

Yu, C., 2005, 'The emergence of links between lexical acquisition and object categorization: a computational study', *Connection Science*, vol. 17, pp. 381-397.