Concept Acquisition through Linguistic Human-Robot Interaction

Joachim de Greeff   Frédéric Delaunay   Tony Belpaeme

School of Computing and Mathematics
University of Plymouth, UK
Email: joachim.degreeff@plymouth.ac.uk

Abstract— **In this paper we give a brief description of conceptual theories, both from a human and a machine perspective. We discuss recent insights from developmental studies on the influence of language and other social factors on concept learning. Based on this we describe a model which incorporates some important aspects and is able to support concept learning trough linguistic interaction. We also describe the prototype of a robotic setup on which our conceptual model will be implemented and which can learn conceptual knowledge from a human teacher through linguistic interaction.**

I. INTRODUCTION

In order to behave properly in a dynamical, ever changing environment and interact with humans, a mobile robot should be equipped with adequate cognitive abilities. Arguably, quite important is the ability to handle everyday concepts. Humans display a capacity to talk about, reason over, infer and generate new concepts with remarkable ease. For a robotic system however, to cope with this is far from trivial. This is in part due to the imprecise nature of many everyday concepts. Hence, in order to properly implement concept handling within an artificial system it makes sense to draw inspiration from theories about how humans deal with concepts. After all, the aim is to gear up a robotic system for interaction with humans.

Concepts have been the topic of investigation for quite some time within cognitive science and related disciplines. Many models and theories have been put forward. In this paper we give a brief overview of conceptual theories and describe a model implementing concept learning strategies. There is evidence that language and interaction plays a crucial role in the formation of concepts in humans, and hence we want to explore how aspects of this can be applied in artificial concept formation. To do so, we propose a conceptual learning model in which language is a prime factor for the acquisition of new concepts and which incorporates some properties deemed to be important within conceptual literature.

II. CONCEPTUAL THEORIES: HUMAN AND MACHINE PERSPECTIVE

Regarding theories about concepts different perspectives can be taken. On the one hand there is the human/psychological perspective, in which theories are postulated as to explain how human cognition handles concepts. On the other hand there is the machine learning perspective, in which conceptual theories describe how artificial systems may handle the formation, representation and use of concepts. Both perspectives will be discussed in more detail.

Within the literature the terms 'category' and 'concept' are frequently used interchangeably[1]. In [1] (p.62) several definitions of 'concept' and 'category' are discussed. The main conclusion is that the terms 'concept' and 'category' do not have a clear definition and need to be explicitly defined within the theoretical framework in which they are used. However, as Murphy notes in [2] (p.5), "being to fussy about saying *concept* or *category* leads to long-winded or repetitious prose (...) with little advantage in clarity". Henceforth, in this paper we will use the terms 'concept' and 'category' loosely as synonyms, unless specific differentiation is required for reasons of clarity.

A. The human perspective

There is a range of different theoretical stances on human concepts and conceptualisation. The classical theory essentially regards a concept as a list of binary necessary and sufficient properties which an object needs to hold in order to belong to a specific concept. So BACHELOR may be comprised of the properties "male", "is not married" and "adult", and consequently, every object that holds these properties is a BACHELOR. This classical view was dominant until the 1950's when more and more drawbacks were raised, most notably the observation by Rosch [3] that many everyday concepts are prototypical in nature, rather than logical defined, i.e. humans judge certain instances of a specific concept to be more typical than others. For example, for the concept BIRD, the instance ROBIN is thought to be more "bird-like" than the instance PENGUIN. Hence, it seems that instances of a concept exhibit a graded membership to an idealised prototype. Since then the theory about concepts has developed substantially, notable through postulating the "exemplar view" [4] in which a concept is seen as a collection of exemplars rather than an idealised prototype, the "theory-theory" or "knowledge approach" [5] which regards concepts as a part of mental theories/general knowledge a person holds about the world and states that learning a concept always happens in interaction with this

[1] As are 'categorisation', 'conceptualisation', 'classification' and 'discrimination'; all of these terms more or less refer to an act of assigning some kind of input data as belonging to some kind of group.

general knowledge, the “neoclassical theory” [6], [7], which strives to rescue the classical view by modifying problematic aspects and “conceptual atomism” [8] which views every concept as atomic and not holding any structure of properties and relations (for an extensive overview see [9] and [2]).

B. The machine perspective

Within machine concept learning and handling, a number of different approaches can be distinguished. Quite a large body of research is done within machine vision. In this, concept learning typically consists of analysing a visual stimulus and identifying key regions related to concepts. The emphasis seems usually not to be so much on the build-up of a conceptual knowledge structure, but rather on the ability to categorise incoming data into classes. Typically, models capable to do so need to be trained first in either supervised or unsupervised fashion, which is in contrast with the “fast mapping” phenomenon observed in children [10]. As opposed to human conceptual models however, language is usually not taken into account in the more straightforward categorisation tasks[2].

Within machine approaches to concept learning, categorisation of stimuli can roughly be divided into two different methods. Stimuli are either assigned to a certain class based on some kind of rule, or based on some kind of distance measurement to exemplar members of a class. Neural network approaches fall somewhat outside this distinction, because their distributed fashion makes it not straightforward to identify either rules or explicit distance calculations.

Rule-based selection can be seen as the machine version of the “knowledge approach”. In this, the learning algorithm maintains a set of rules or hypotheses about what would constitute a category and tries to classify incoming stimuli according to these rules. In general, this requires at least some knowledge about the domain and the type of categories to be expected, but some of these models could start out with very general knowledge and adapt their concept hypotheses based on encountered exemplars [11], [12]. This relates to the field of Expert Systems and Case Base Reasoning [13], although here the emphasis is generally more on applying the right expert knowledge to the proper situations.

On the other hand, within similarity selection an incoming stimulus is compared to an existing body of conceptual knowledge. In effect, it is measured how ‘similar’ the stimulus is to known concepts through calculation of the distance. In order to do so, the dimensions of existing concepts and incoming stimuli need to be compatible. An existing concept may either consist of a set of exemplars which together comprise the concept, analogue to the “exemplar view” from the human perspective, or the concept may be represented by an idealised prototype (prototype theory) which is typically derived from observed exemplars. The rule-based versus similarity selection is a classical distinction, which can be viewed as analogue to the concept debate in the human perspective. More modern theories however do not always draw the line so explicitly. For instance, Smith et. al [14] argued that categorisation may exhibit both rule-based and similarity selection features.

Neural network based architectures have also been proposed as candidates for storing and classifying knowledge. For instance, McClelland and Rumelhart [15] describe a model which is able to learn both general and specific knowledge and is able to classify in both prototypical and rule-based manner. Another well known example of connectionist approach to concept learning is ALCOVE of Kruschke [16], which is based on Nosofsky’s generalized context model [17], both using the exemplar approach. Overall, there is a vast amount of literature on classification using neural networks, but a detailed discussion of this is beyond the scope of this paper. For an overview, see [18]. Notably though, most of neural network categorisation research is aimed at classification of stimuli, and not so much at more general conceptual aspects or the development of psychological plausible conceptual models.

Probabilistic models have also been taken into account; both rule-based and similarity based approaches as described above may incorporate probabilistic aspects, but probabilistic models have also been studied exclusively. Tenenbaum ([19], [20]) for instance, proposed a Baysian framework for concept learning, which incorporates both rule-based and similarity-based selection mechanisms. Within such a Baysian framework it is possible to combine prior knowledge or constrains (like the ones proposed by Markman [21]) with observed statistical exemplar data.

III. Concept Learning in Children

Within development psychology a lot of research has been done on how children learn the meaning of words and/or new concepts. Some different aspects that influence learning can be identified. These are: 1) the role of language, 2) interaction and 3) learning constrains. We will discuss these aspects in more detail.

A. The role of language in concept learning

The distinction between learning concepts (internal representations of a certain worldly class) and learning the words describing these concepts is not always made clear in the literature. Quite often this is seen as one process, in which learning a new word entails learning a new concept. In this paper however, we view these two as being separate aspects, which are nevertheless very much intertwined. Indeed, in recent years it has been acknowledged by several authors that the formation of new categories is heavily influenced by language. As such, the labels used to describe a stimulus representing a new concept govern the way in which this new data will be integrated into existing conceptual knowledge. Several studies have shown that young children, in addition to learning directly from sensory exploration, rely on linguistic labels to acquire the meaning of words (for an

[2]This is not to say that language is always taken into account within theories of human concepts, which is definitely not the case. As language is very much tied in with general cognition however, the link between language and concepts is usually acknowledged within human conceptual theories, see section III-A.

overview of early language acquisition see [22]). Xu [23], for example, demonstrated how linguistic labels help 9-month old infants to establish a representation for different objects. Learning without linguistic labels, or with the presence of tones, sounds or emotional expressions is not effective. This implies that language is crucial in acquiring novel concepts from a very early age on. Plunket et al. [24] came to the same conclusion in a tightly controlled experiment in which they demonstrated how category formation in 10-month old infants is influenced by linguistic labels. Linguistic labels also have an effect on category learning in adults; adults who learn a new category did so significantly faster and showed more robust category recall when the learning experience was accompanied by novel linguistic labels ([25]; [26]). Aforementioned studies show that linguistic labels facilitate category acquisition, both in pre-linguistic infants and adults. These insights tie in with linguistic relativism, which states that language and cognition influence each other. Recently, linguistic relativism gained renewed attention as a series of psychological experiments demonstrated how perception of stimuli and use of categories is influenced by the words we know; this has been notably demonstrated for categories of time, colour and space (e.g. [27]; [28]; [29]).

B. Interaction

In developmental psychology it has been known for a while that social and affective interaction is central to language development and, by extent, to concept acquisition. Young learners and their caretakers engage in *intersubjectivity* [30], the common denominator for interactions involving the learner's understanding of emotion and thought. There is evidence that the acquisition of language and specifically of vocabulary requires the young child to interact with a human caretaker: Grela, Krcmar and Lin [31] show how 15 to 21 month old children benefit from interaction: learning novel words was significantly more efficient when joint attention and interaction was involved. This was contrasted to learning experiences presented on a television screen, thus not involving joint attention or interaction; results showed how televised learning experiences were up to half as effective as actual interaction (interestingly, the study suggested that children are least likely to learn novel words presented by animated characters on television).

Baldwin and Moses [32] describe a large body of evidence indicating that children need social understanding to properly learn a word-object mapping. That is, temporal contiguity of a word with an object is often not enough for children to accept the word as a proper label, additional social cues indicating that a caregiver is indeed referring to the specific object is required to properly learn new words. These social clues could be joint attention ([33]; [34]) and/or non-verbal communication like facial expressions or gestures.

The importance of interaction for learning has also been acknowledged within the human-robot interaction community (e.g. [35]). Several researchers have examined the role of interaction by building robotic models that learn from the environment through interacting with it, rather than plain observation. Breazeal and Scassellati [36] for instance, describe a system which allows a robot to regulate the level of interaction, so that it gets neither too much nor to little stimulation from it's surroundings. Thus, the system actively creates an optimal learning environment for itself. Also in [37] human to robot teaching was studied, revealing the importance of mutual feedback to keep the interaction active, that is to say: focused. Joint attention within robotic models was investigated by Kaplan en Hafner, in [38] they tried to identify the underlying skills necessary to achieve functional joint attention between a human and a robot or between two robots.

C. Concept learning constrains

The well known example of Quine [39] about the indeterminacy of reference led scholars to argue that children use certain constrains that greatly facilitate the faced induction problem when learning new concepts. These constrains are: the whole object assumption: a new concept label refers to the whole object that is perceived, rather than to parts of it. Taxonomic assumption: it is assumed that the new concept exists within a taxonomy, with possibly a bias to basic level concepts. Mutual exclusive assumption: a label refers to one concept. Because of this children do not need any negative examples; being shown what a certain concept is implies also what it is not. Related to the whole object assumption is the shape bias, which seems to cause young children to pay attention to the shape of the object when confronted with a new stimulus [40]. In response to this, in [20] it is argued that using these constrains "would solve the problem of induction for one important class of words (concepts, ed.), but at the cost of making the rest of the language unlearnable". Tenenbaum and Xu propose a model based on Bayesian inference which is able to accommodate milder forms of the constrains described above in combination with the statistical structure of observed exemplars.

IV. Our model

We developed a computational model which is able to learn new concepts through linguistic interaction with a teacher. As described in section III, there exist a large body of evidence suggesting that language is a prime driver for the formation of new concepts and that interaction is important for proper learning. Because of this, our aim is to incorporate such learning dynamics within the model. To do so, we adapted learning interaction based on Language Games ([41]; [42]) to a teacher-learner scenario. To represent conceptual knowledge, we use a Conceptual Space [43], which is a geometrical representation allowing for similarity measurement. Hence, it falls in the similarity selection approach, although it is not our aim to present the model as a champion for this view. Rather, we believe that such a representation is capable of displaying much of the characteristics of human conceptualisation, as well as being intuitively appealing because of the possibility of geometrical display with meaningful dimensions. The learning through language interaction and the CS will be described in more detail below. The model is designed to be implemented on a robotic system which can learn concepts through interaction with a human (or a robotic) teacher. It is based on previous work as reported in [44], in which we studied the effect of adding interactive features to the learning interaction.

A. *Learning trough language interaction*

In the model, the agent is able to learn new concepts through engaging in a language game (LG) with a teacher. A LG is implemented as a combination of a discrimination game and a guessing game ([41]; [42]) and typically consist of two agents that try to establish a shared lexical description for a certain observation. One agent acts a teacher (and hence should have some a priori conceptual knowledge) and the other as a learner. Both agents are given a set of training stimuli, called the context, from which one specific stimulus is assigned as topic which is known by the teacher only. The teacher then communicates the topic to the learner by stating its associated linguistic label. If the learner is able to correctly identify the topic from the context through the perceived label, the game succeeds. If the learner is not able to correctly identify the topic, either because the label is not known by the learner or because the learner points to the wrong item from the context, the game fails. A failed game provides an opportunity for the learner to improve its conceptual and lexical knowledge. By employing a sequence of language games, the learner is able to build a body of knowledge containing conceptual prototypes and associated labels.

B. *Knowledge representation*

In order to store conceptual knowledge we use a Conceptual Space (CS). A CS consists of a geometrical representation of conceptual knowledge along various quality dimensions. In a nutshell, a conceptual space is a collection of one or more domains (like colour or tone), where a domain is postulated as a collection of inseparable quality dimensions with a metric. Examples of quality dimensions are *weight*, *temperature*, *brightness*, *pitch*, *loudness* and *RGB values*. For instance, to express a colour in RGB encoding[3] the different quality dimensions 'red', 'green' and 'blue' are all necessary to express colour values and are hence inseparable. Other domains may consist of more or just one quality dimension. In its simplest form, a concept can be represented as a point in the conceptual space, where the coordinates of the point determine the features of the concept. For example, the concept RED is represented as a point on (255, 0, 0) in the RGB colour domain and the concept BLUE as a point on (0, 0, 255). In principle any domain may be used, although for some domains it might be easier to extract the relevant dimensions than for others.

A newly observed stimulus can be classified as belonging to a particular existing colour concept by calculating the weighted distance from the stimulus to every concept already present in the conceptual space. The observed stimulus is then assigned to the closest existing concept. Furthermore, a CS allows for the representation of concepts through prototypes, which enables it to display typicality effects observed in human conceptualisation. A conceptual prototype is built through the addition of exemplars for the specific concept, where the mean values of all dimensions encode for the coordinates of the prototype, and the variance of all exemplars determines the prototype's size. A conceptual prototype will therefore not be an exact point in the conceptual space, but rather define a certain region. However, in order to calculate the coordinates and variance of a prototype when a new exemplar is added properly, all exemplars comprising the prototype are maintained within memory. Through distance calculation from exemplar to prototype, it is possible to derive the most prominent exemplars for a certain concept. Thus, the model exhibits both prototype and exemplar properties.

C. *Lexicalisation*

The conceptual knowledge represented in the CS is tightly linked to a lexicon of linguistic labels used to describe the concepts. The linguistic labels can potentially be stored as a string of characters, a visual icon or an acoustic sample. Moreover, these labels play a crucial role in the formation of new concepts. Typically, a learning agent is confronted with a stimulus accompanied by a linguistic label. The stimulus is then integrated as exemplar data into the existing conceptual knowledge of the agent. The accompanying label may influence to which existing concept the new data should be assigned to. And the other way around, the meaning of the linguistic labels can be perceptually grounded through the values of the associated concepts. Concepts are linked with labels through an association matrix which determines the strength of the connection between every known concept in the CS and every label in the lexicon. Hence, when the agent needs a label to express a specific concept or vice versa, this can be found through consulting the association matrix.

D. *Properties of the model*

The model exhibits a range of properties consistent with findings in human concept literature[4]:

Learning through linguistic interaction. As described, by using language games as a learning mechanism, the linguistic labels which accompany stimuli play a crucial role in integration of the stimuli into existing conceptual knowledge. As in [24], the label offered by the teacher as a description of an unknown stimulus determines to which concept the stimulus will be assigned to. Thus, describing two distinct stimuli with two different labels results in the creation of two

[3]RGB is a technical colour representation suited for reproducing colour on display devices. Red, Green and Blue values here range from 0 to 255. Other more psychologically plausible colour representations exist, such as CIE L*a*b* which is aimed at modelling human colour appearance.

[4]It is not our aim to develop a full-fledged model able to accommodate all effects observed in human concept handling. Rather, we wish to buil a functional model which can perform "well enough", i.e. which implements certain features deemed to be important. The main emphasis is on the fact that language is an important driver for the formation of concepts and that the model can be applied in a human-robot interaction scenario.

different concepts (provided the stimuli cannot be linked to existing concepts of the learning agent), and describing the same stimuli with only one label results in the creation of a more general concept, comprised of both stimuli.

Typicality effects. Due to the prototypical nature of concept storage a stimulus can be graded as belonging to a specific concept, based on measuring the distance. Distance d between i and j is measured using the weighted Minkowski r-metric:

$$d_{ij} = \left(\sum_{k=1}^{N} w_k |x_{ik} - x_{jk}|^r \right)^{\frac{1}{r}} \tag{1}$$

where r denotes the type of metric ($r = 1$ is the Manhattan distance and $r = 2$ is the Euclidean distance) and w the weight of the dimension, determining how much emphasis should be put on it. To do justice to psychological evidence of how humans view concepts ([45]; [17]), the distance can be converted into a similarity measurement. Similarity s between i and j is computed as an exponentially decaying function of distance:

$$s_{ij} = e^{-cd_{ij}} \tag{2}$$

where c is a "sensitivity" parameter.

Interactive features. A learning agent can make use of some interactive learning features, which allow it to actively influence the learning process. These interactive features are implemented within the language game interaction and are designed to help the learning agent to develop conceptual knowledge more quickly, while keeping psychological relevance in mind. There are three interactive features: active learning, knowledge querying and contrastive learning. Active learning gives the learner a bias to learning unknown stimuli and relates to novelty preference which is typically observed in young children, knowledge querying helps the learning agent to solidify its conceptual knowledge by testing uncertain label-concept mappings with the teacher and contrastive learning allows the learner to use certain stimuli as negative examples, which bears similarities to lateral inhibition [46] and lexical contrast [47].

E. Preliminary results

In a series of simulation experiments we tested our model on the influence of the interactive features (as described above) on learning. Normal language game interaction on the learning of basic English colour names was used as a base setting and this was compared the performance of agents with interactive features enabled. A teaching agent with preloaded knowledge about colour concepts engaged in language game interaction with a learning agent with blank conceptual knowledge. Performance of the learning agent was measured by calculating the percentage of correctly named test stimuli[5] after each language game. The number of training interactions the agents engaged in was 2000, and the experiment was replicated 300 times. As can be observed in Figure 1, the mean performance of an agent with all interactive features enabled is a marginally but statistically significant better than a normal learning agent (Welsch two sample t-test; $p < 0.0001$). For a more extensive description, see [44].

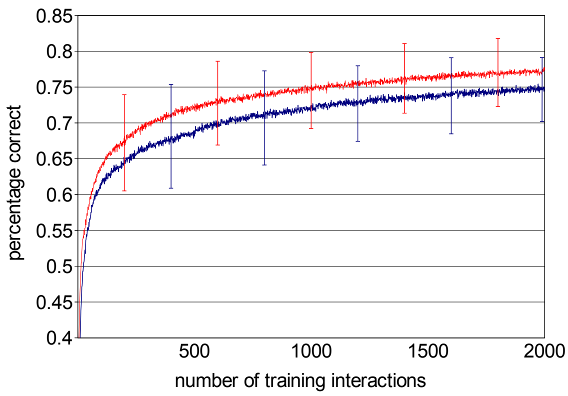

Fig. 1. Performance of a normal agent (blue, bottom) versus an interactive agent (red, top) over 2000 training sessions.

F. Extention proposal

In order to increase the psychological plausibility of the model, we are currently working on an extension of the current setup which allows for two additional features deemed to be important for conceptual modelling. These features are compositionality and hierarchy. We will discuss these in more detail below.

Compositionality. Compositionality refers to the the ability to build new, complex concepts from existing, more simple, concepts. The classical approach is to take the intersection of the two concepts to be combined; thus STRIPED APPLE would be all objects exhibiting the properties STRIPED and APPLE. This yields to problems however, because it is impossible to take the intersection for complex concepts like PINK ELEPHANT and STONE LION, which are nevertheless very well handled by humans. There exist multiple approaches in the literature of how to model this. And example is the selective modification model [48], see [49] for some comparisons. We opt for an implementation following Gärdenfors [50]. In this, analogue to the English language, the order of the two concepts to be combined is important. Thus, when combining the concepts X and Y to XY, X acts as a modifier of Y, i.e. RED BRICK describes some kind of BRICK and BRICK RED describes some kind of RED. In effect, the properties of X (i.e. the quality dimensions) corresponding to properties of Y replace these properties of Y. If Y does not have the properties of X, these are simply added, if Y does have them, they are overruled by the properties of X. X may also influence certain other properties of Y. For instance, the addition of WOODEN (X) to SPOON (Y) may yield to revising the 'size' property of Y (making it bigger), or block certain properties as in STONE (X) LION (Y), where X blocks properties of Y like 'living' and

[5]Both teacher and learner are confronted with 100 colour stimuli drawn randomly from a set of 25,000. If the learner provides the same colour label as the teacher does for a given stimulus, it scores 1, otherwise 0.

'sound' and 'habitat' as they do not apply to STONE. More specifically, when combining X and Y, X may entail certain properties that may influence Y. Y however, determines the context in which these properties of X are applied (this is called a 'contrast class' by Gärdenfors).

Hierarchy. Typically, conceptual ontologies are seen as a taxonomic structure in which subordinate concepts inherit properties from their superordinate counterparts. In order to obtain such a structure within the initial non-connected individual concepts residing in the CS, we plan to apply a Formal Concept Analysis (FCA) [51] algorithm on the data in the CS. Original developed as a sub-field of applied mathematics, FCA allows for the "mathematisation" of concepts by postulating them as a hierarchically ordered lattice of objects and properties. FCA can be applied to a whole range of fields, including cognitive science, linguistics, data mining and economics, see [52] for an overview. Recently, in [53] it was shown that hierarchies constructed through FCA may have some biological basis. In FCA a tripled $\{O, P, I\}$ (called the context) is considered, where O is a non-empty set of objects, P is an non-empty set of properties and I is a binary relation between O and P indicating whether or not object $o \in O$ has property $p \in P$. A (formal) concept is then a pair (A, B), with $A \in O$ and $B \in P$ such that A is the maximal set in O that shares all properties of B and B is the maximal set of properties shared by all the objects in A. The collection of all formal concepts of $\{O, P, I\}$ can be ordered in a hierarchical set-theoretic structure, called the concept lattice, through requiring that all objects of subordinate concepts are also objects in their superordinate concepts. This yields to a formally sound order which is also intuitively easy to grasp. By applying FCA methods as described above, the aim is to have the model store conceptual knowledge in hierarchical fashion.

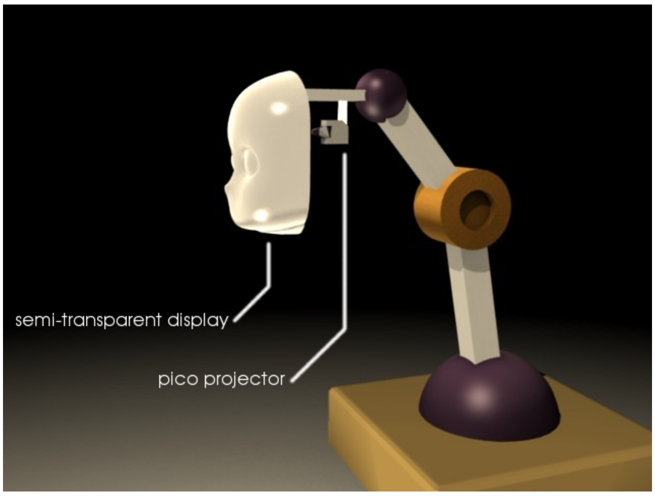

Fig. 2. Sketch of the robotic setup, the image is retro-projected into the opal face mask. The face is mounted onto an articulated neck. The current prototype described in this paper uses a ViewSonic XGA projector, however a picoprojector is to be used in a later version.

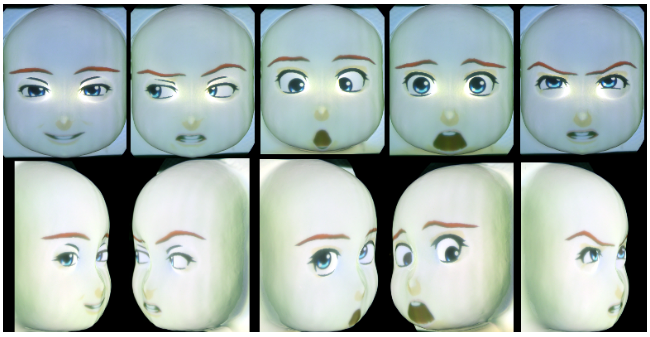

Fig. 3. Faces displaying basic five expressions. From left to right, both top and bottom: happiness, disgust, surprise, fear, anger. The neutral expression is not shown.

G. Human-robot interaction and robot-robot interaction

The conceptual model will be applied within a robotic setup, which will feature Human-Robot interaction (HRI). The robot consists of an articulated neck and attached face, which is able to display a range of emotions and incorporate personality traits, see [54] for an extensive description of the prototype. The face of the robot consists of a projected image, allowing full control of the facial expressions that the robot displays. Figure 2 displays a schematic setup and Figure 3 shows different emotions the robot is able to display through face projection.

With the conceptual model acting as a knowledge representation structure, the robot will interact with human teachers by using various methods like shared attention, speech recognition, facial expression, display of emotions etc. in order to learn conceptual knowledge. Through this setup we plan to study to what extend these aspects influence the effectiveness of HRI. A typical learning scenario will proceed as follows: 1) the robot tries to get the attention from the human, 2) through a (probably predefined) script the robot invites the human to teach, 3) the human shows the robot various items and provides a name (linguistic label) for them, 4) based on its sensor capacities the robot perceives the most relevant data of the objects (colour, shape, other features), 5) based on the given label, the input data is integrated with existing conceptual knowledge, 6) the robot may test its existing knowledge by querying it with the teacher (see interactive features in section III-B). These HRI sessions will allow us to study both the effectiveness of the model as a conceptual knowledge structure, and the effectiveness of the robotic setup with respect to emotions and personality traits.

In a fashion analogue to the HRI described above, the interaction and learning algorithms of the model will be applied to robot-robot learning as well. From a learning robot point of view, there will be no difference if its teacher is human or robot. However, the interaction will be different because it will not be based on a worldly setting in which the robot needs to extract the proper features from the objects. Rather, learning interaction will happen over a network between to software agents exhibiting the conceptual model, with one agent that has some conceptual knowledge acting as a teacher, and another agent acting as a learner.

V. CONCLUSIONS

In recent years it has been shown how young children rely on social and linguistic interaction to bootstrap their conceptual knowledge. Replicating this process onto a robot is important in several ways. Firstly, it uses social and linguistic interaction to reduce ambiguity in the learning process: non-verbal feedback and linguistic labels help to reduce the search space, much in the same way as supervised learning converges faster than unsupervised learning. Secondly, social interaction and language comes naturally to people. Caretakers willingly engage in tutelage, employing a range of communicative cues to structure the learning for young children.

This paper reports design specifications and progress in constructing the learning mechanisms for such a robot and describes how new technology in facial animation on a social robot will be used to engage in social interaction supporting tutelage. More specifically, we have described the characteristics of a conceptual model which incorporates aspects of of concept learning and storage that are considered to be important within conceptual and child developmental literature.

VI. ACKNOWLEDGMENTS

This work is supported by the UK EPSRC CONCEPT project (EPSRC EP/G008353/1) and by the EU FP7 ITALK project (ICT-214668).

REFERENCES

[1] P. Davidsson, "Autonomous agents and the concept of concepts," Ph.D. dissertation, Department of Computer Science, Lund University, 1996.

[2] G. L. Murphy, *The Big Book of Concepts*. MIT Press, 2002.

[3] E. H. Rosch, "Natural categories," *Cognitive Psychology*, vol. 4, no. 3, pp. 328 – 350, 1973.

[4] D. L. Medin and M. M. Schaffer, "Context theory of classification learning," *Psychological review*, vol. 85, pp. 207–238, 1978.

[5] G. L. Murphy and D. L. Medin, "The role of theories in conceptual coherence," *Psychological Review*, vol. 92, no. 3, pp. 289–316, 1985.

[6] R. S. Jackendoff, *Semantics And Cognition*. Cambridge: Mit Press, 1983.

[7] S. Pinker, *Learnability and Cognition: The acquisition of Argument Structure*. Cambridge, Mass.: MIT Press, 1989. [Online]. Available: http://www.isrl.uiuc.edu/ amag/langev/paper/pinker89book.html

[8] J. A. Fodor, "Information and representation," in *Information, language, and cognition*, P. Hanson, Ed. Vancouver, BC: University of British Columbia Press, 1990, pp. 175–190.

[9] E. Margolis and S. Laurence, *Concepts: Core Readings*. MIT Press, 1999.

[10] E. V. Clark, "What's in a word? on the child's acquisition of semantics in his first language," in *Cognitive development and the acquisition of language*, T. E. Moore, Ed. New York: Academic Press, 1973, pp. 65–110.

[11] M. J. Pazzani, "Influence of prior knowledge on concept acquisition: Experimental and computational results," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 17, no. 3, pp. 416–432, 1991. [Online]. Available: http://citeseer.ist.psu.edu/pazzani91influence.html

[12] R. Mooney, "Integrating theory and data in category learning," in *Categorization by Humans and Machines*, G. V. Nakamura, D. L. Medin, and R. Taraban, Eds. Academic Press, 1993, vol. 29, pp. 189–218.

[13] A. Aamodt and E. Plaza, "Case-based reasoning: foundational issues, methodological variations, and system approaches," *AI Commun.*, vol. 7, no. 1, pp. 39–59, 1994.

[14] E. E. Smith, A. L. Patalano, and J. Jonides, "Alternative strategies of categorization," *Cognition*, vol. 65, no. 2-3, pp. 167–196, 1998.

[15] J. L. McClelland and D. E. Rumelhart, "Distributed memory and the representation of general and specific information," *Journal of Experimental Psychology: General*, no. 141, pp. 159–197, 1985.

[16] J. K. Kruschke, "Alcove: an exemplar-based connectionist model of category learning." *Psychological review*, vol. 99, no. 1, pp. 22–44, January 1992.

[17] R. M. Nosofsky, "Attention, similarity, and the identification-categorization relationship," *Journal of Experimental Psychology-General*, vol. 115, no. 1, pp. 39–57, 1986.

[18] G. Zhang, "Neural networks for classification: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 30, no. 4, pp. 451–462, Nov 2000.

[19] J. B. Tenenbaum, "A bayesian framework for concept learning," Ph.D. dissertation, MIT, 1999.

[20] J. B. Tenenbaum and F. Xu, "Word learning as bayesian inference," in *In Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Erlbaum, 2000, pp. 517–522.

[21] E. Markman, *Categorization in children: Problems of induction*, ser. Bradford Books. Cambridge, MA: MIT Press, 1989.

[22] P. Bloom, *How children learn the meanings of words*. Cambridge, MA: the MIT Press, 2000.

[23] F. Xu, "The role of language in acquiring object kind concepts in infancy," *Cognition*, vol. 85, no. 3, pp. 223 – 250, 2002.

[24] K. Plunkett, J. Hu, and L. B. Cohen, "Labels can override perceptual categories in early infancy," *Cognition*, vol. 106, no. 2, pp. 665–681, February 2008.

[25] G. Lupyan, "Labels facilitate learning of novel categories," in *Proceedings of the 6th International Conference on the Evolution of Language*, 2006, pp. 190–197.

[26] G. Lupyan, D. H. Rakison, and J. L. McClelland, "Language is not just for talking - redundant labels facilitate learning of novel categories," *Psychological Science*, vol. 18, no. 12, pp. 1077–1083, 2007.

[27] A. Gilbert, T. Regier, P. Kay, and R. Ivry, "Whorf hypothesis is supported in the right visual field but not the left," in *Proceedings of the National Academy of Sciences*, 103(2) 2006, pp. 489–494.

[28] A. Majid, M. Bowerman, S. Kita, D. B. M. Haun, and S. C. Levinson, "Can language restructure cognition? the case for space," *Trends in Cognitive Sciences*, vol. 8, no. 3, pp. 108 – 114, 2004.

[29] D. Roberson, H. Pak, and J. R. Hanley, "Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from korean," *Cognition*, vol. 107, no. 2, pp. 752 – 762, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/B6T24-4PVY2XR-1/2/3b1d0525225794ef4c5d19956a236564

[30] C. Trevarthen, "The concept and foundations of infant intersubjectivity," in *Intersubjective Communication and Emotion in Early Ontogeny*, S. Braten, Ed. Cambridge, UK: Cambridge University Press, 1998, pp. 15–46.

[31] B. Grela, M. Krcmar, and Y. Lin, "Can television help toddlers acquire new words?" 2004.

[32] D. Baldwin and L. Moses, "Links between social understanding and early word learning: Challenges to current accounts," *Social Development*, vol. 10, pp. 309–329, 2001.

[33] D. Baldwin, "Understanding the link between joint attention and language," in *Joint attention Its origins and role in development*, C. Moore and P. Dunham, Eds. Hillsdale, NJ: Lawrence Erlbaum, 1995, p. 131158.

[34] M. Tomasello, "Joint attention as social cognition," in *Joint attention Its origins and role in development*, C. Moore and P. Dunham, Eds. Hillsdale, NJ: Lawrence Erlbaum, 1995, p. 103130.

[35] J. Cassell and A. Tartaro, "Intersubjectivity in human-agent interaction," *Interaction Studies*, vol. 8, no. 3, pp. 391–410, 2007.

[36] C. Breazeal and B. Scassellati, "Infant-like social interactions between a robot and a human caregiver," *Adapt. Behav.*, vol. 8, no. 1, pp. 49–74, 2000.

[37] A. L. Thomaz, "Socially guided machine learning," Ph.D. dissertation, MIT, 2006.

[38] F. Kaplan and V. Hafner, "The challenges of joint attention," in *Proceedings of the 4th International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic System*, L. Berthouze, H. Kozima, C. Prince, G. Sandini, G. Stojanov, G. Metta, and C. Balkenius, Eds. Lund University Cognitive Studies 117, 2004, pp. 67–74.

[39] W. V. O. Quine, *Word and Object*. Cambridge, MA: The MIT Press, 1960.

[40] B. Landau, L. B. Smith, and S. S. Jones, "The importance of shape in early lexical learning," *Cognitive Development*, no. 3, pp. 299–321, 1988.

[41] L. Steels and T. Belpaeme, "Coordinating perceptually grounded categories through language: A case study for colour," *Behavioral and Brain Sciences*, vol. 28, no. 4, pp. 469–89, August 2005, target Paper, discussion 489-529.

[42] T. Belpaeme and J. Bleys, "Explaining universal colour categories through a constrained acquisition process," *Adaptive Behavior*, vol. 13, no. 4, pp. 293–310, 2005.

[43] P. Gärdenfors, *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: the MIT Press, 2000.

[44] J. De Greeff, F. Delaunay, and T. Belpaeme, "Human-robot interaction in concept acquisition: a computational model," in *Proceedings of the International Conference on Development and Learning (ICDL)*, Shanghai, submitted, 2009.

[45] R. Shepard, "Toward a universal law of generalization for psychological science," *Science*, vol. 237, no. 4820, pp. 1317–1323, 1987.

[46] M. Oliphant, "The learning barrier: Moving from innate to learned systems of communication," *Adaptive Behavior*, vol. 7, no. 3/4, 1999. [Online]. Available: http://www.isrl.uiuc.edu/ amag/langev/paper/oliphant99theLearning.html

[47] E. V. Clark, *The lexicon in acquisition*. Cambridge: Cambridge University Press, 1993.

[48] E. E. Smith, D. N. Osherson, L. J. Rips, and M. Keane, "Combining prototypes: A selective modification model," *Cognitive Science*, vol. 12, no. 4, pp. 485–527, 1988.

[49] E. Machery and L. Lederer, "Simple heuristics for concept combination," under review, in M. Werning, W. Hinzen, and E. Machery (Eds.), The Oxford Handbook of Compositionality.

[50] P. Gärdenfors, "Concept combination: a geometrical model," in *Logic, Language and Computation*, L. Cavedon, P. Blackburn, N. Braisby, and A. Shimojima, Eds. Stanford, CA: CSLI Publications, 2000, vol. 3, pp. 129–146.

[51] R. Wille, "Restructuring lattice theory: an approach based on hierarchies of concepts," in *Ordered Sets*, I. Rival, Ed. Dordrecht: Proc. Nato Advanced Study Institute, vol. 1. Reidel, 1982, pp. 445–470.

[52] B. Ganter, G. Stumme, and R. Wille, Eds., *Formal Concept Analysis: Foundations and Applications*, ser. Lecture Notes in Artificial Intelligence, vol. 3626. Heidelberg: Springer, 2005.

[53] T. Wennekers, "On the natural hierarchical composition of cliques in cell assemblies," *Journal of Cognitive Computation*, 2009.

[54] F. Delaunay, J. De Greeff, and T. Belpaeme, "Towards retro-projected robot faces: an alternative to mechatronic and android faces," in *Proceedings of the International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Toyama, submitted, 2009.