PALADYN Journal of Behavioral Robotics

VERSITA

Research Article · DOI: 10.2478/s13230-013-0105-9 JBR · 3(4) · 2012 · 200-208

Modelling Concept Prototype Competencies using a Developmental Memory Model

Paul Baxter[1]*, **Joachim de Greeff**[1], **Rachel Wood**[2], **Tony Belpaeme**[1]

1 Centre for Robotics and Neural Systems, Cognition Institute, Plymouth University, U.K.

2 Intelligent Computer Systems, Faculty of Information & Communication Technology, University of Malta, Malta

Received 16-12-2012

Accepted 27-03-2013

Abstract

The use of concepts is fundamental to human-level cognition, but there remain a number of open questions as to the structures supporting this competence. Specifically, it has been shown that humans use concept prototypes, a flexible means of representing concepts such that it can be used both for categorisation and for similarity judgements. In the context of autonomous robotic agents, the processes by which such concept functionality could be acquired would be particularly useful, enabling flexible knowledge representation and application. This paper seeks to explore this issue of autonomous concept acquisition. By applying a set of structural and operational principles, that support a wide range of cognitive competencies, within a developmental framework, the intention is to explicitly embed the development of concepts into a wider framework of cognitive processing. Comparison with a benchmark concept modelling system shows that the proposed approach can account for a number of features, namely concept-based classification, and its extension to prototype-like functionality.

Keywords

Cognitive Architecture · Concept Development · Conceptual Spaces · DAIM · Distributed Associative Memory

1. Introduction

From the perspective of autonomous robotic agent functionality, humans form an important source of inspiration as they provide a working example of desirable behaviours. Theory and data are therefore drawn from the empirical sciences (psychology, neuroscience, etc) in the search for aspects of functionality and mechanisms that can be applied to computational systems. Furthermore, from the perspective of the design of synthetic agents for the purpose of human-robot interaction (particularly social), it is desirable that the functionality of these agents reflects human cognitive processing, e.g. [6, 22]. This is to ensure that the humans' expectations of the system may be fulfilled, and that the system may better assess the behaviour (be it actual or expected) of the human. In this context, it is not just desirable to achieve the highest possible efficiency for the robotic agent, it is also desirable to be able to account for the sources of non-optimal (in the sense of accuracy) performance in humans. Finally, in the context of autonomous operation, it is also desirable that robotic agents acquire the desired competencies in a manner that is not purely reliant on supervised learning.

One particularly important aspect of human behaviour is the ability to deal with conceptual knowledge. It is a fundamental requirement for human cognition, and so is likewise a necessary competence for autonomous synthetic agents [8, 19]. Concepts have long been regarded as logical definitions, effectively specifying a list of necessary and sufficient properties that describe a concept. For example, the concept BACHELOR may be comprised of the properties *adult*, *male* and *is not married*; consequently, everything that is an adult male and is unmarried is therefore a BACHELOR. This definition of concept representation is known as the classical theory [16, 20, 27].

However, fundamental problems with the classical theory of concept have been identified, in particular the fact that for a lot of concepts it is very hard, if not impossible, to come up with a logical set of defining properties. And even if such a set could be identified, it is apparent from psychological studies that people do not adhere to these definitions consistently. This has resulted in the formulation of new theories that closer matched human performance. One such reformulation is that a fundamental characteristic of human concepts is the use of prototypes [24]. This theory postulates a concept as an idealised version constructed from examples that people have experienced throughout their life. For the concept BIRD people have a prototype that represents the ideal bird, and any encounters they have in the real world are matched with this prototypical version. The more similar an observation is to the prototype, the more likely it is to be considered as an instance of this concept. A prototype is thus a summary representation that specifies the properties of the concept, where the properties need not have equal importance. Not all properties are necessary, as in the classical theory, but rather they describe which properties instances of the concept in general tend to posses. The process of identifying an object in the world thus entails a matching to known prototypes.

This perspective has displaced the notion that concepts can be defined through solely logical definitions, as it has been shown that people are readily able to assess an instance of a concept as being *more* or *less* typical of the concept prototype, where this assessment is based on a similarity judgement, rather than a property-checklist matching procedure [24].

We seek to provide an account of how these features of concept utility can be applied to robotic agents. Increasing evidence from the study of the substrate of human cognition indicates that cognitive functionality is overlapping and distributed, operating on a number of common principles, e.g. [1, 15]. In considering one type of cognitive competence,

*E-mail: paul.baxter@plymouth.ac.uk

this evidence thus indicates that there is a need to consider the wider implications of the required functionality in terms of cognitive processing and architecture, e.g. [5, 29], and interaction with the environment, e.g. [23]. This widening of scope underlines the importance of considering ontogeny: exploring the developmental substrate of such integrated functionality provides insight not only into how such cognitive competence arises, but also indicates a means to achieve autonomous operation for robotic agents [31], with the changes in perspective on system design and evaluation that this entails [26].

In this paper, the question of how to account for concept utility embedded in wider cognitive processing within a developmental framework is explored. We propose and apply a system inspired by a neuropsychological perspective on memory: the Distributed, Associative and Interactive Memory (DAIM) model. To act as a benchmark for human performance, a Conceptual Spaces (CS) system is used, whose account of prototypes matches closely that found in human behaviour [10, 24]. Whilst a good predictive model, the CS system is rather static in structure, and it is unclear how conceptual learning over time can be accounted for in an unsupervised manner. The purpose of this comparison is to see how the developmental DAIM system compares with CS; whether DAIM can account for those features of conceptual knowledge processing exhibited by humans. It is thus equally important to both account for correct classification as it is to account for errors in the identification of concepts; this study is not intended as the proposal of an algorithm for optimal classification performance.

After an introduction to the computational models used to explore the issues of concept prototype acquisition in this paper (Section 2), the zoo dataset used for this exploration is described (Section 3.1). The experimental procedure (Section 3.2) and results obtained are subsequently presented (Section 4), demonstrating classification and prototype-like functionality. These are then discussed in the context of concept development in autonomous synthetic systems (Section 5).

2. Architecture of Models

2.1. The Conceptual Spaces (CS) Model

A Conceptual Space (CS) consists of a geometrical representation in vector space along various quality dimensions [10]. This perspective on concept representation is consistent with accounts of human behaviour (e.g. [24]); a CS model is therefore suitable for use as a benchmark system against which the performance of DAIM can be assessed. A CS is a collection of domains (like colour, shape, or tone), where a domain is postulated as a collection of inseparable sensory-based quality dimensions with a metric [11]. For instance, to express a point in the colour domain using an RGB encoding, the different quality dimensions *red*, *green*, and *blue* are all necessary to express a certain colour and are therefore inseparable. 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 (e.g. Figure 1).

Crucial to modelling concepts in a CS is the ability to take a distance measurement. For each of the dimensions involved, a suitable metric to calculate distance between coordinates on this dimension must be defined. For the case of numerous dimensions the Euclidean distance is typically the most appropriate. For example, for colour, a normalised RGB space can be defined, such that the Euclidean distance between any two points in this space can be calculated (Figure 1). The metric can be augmented with a weight (w) to allow certain dimensions to be more prominently expressed than others.

Within a CS the learning of prototypes can be modelled by exposing the model to instances with associated labels. For example, various shades of red could be presented, each with the label 'Red': the prototype in this case could correspond to the mean coordinates of all instances. After this learning phase, the model is able to classify new examples as belonging to some known class, and specify how typical the example is, based on a distance measurement between this example and the various possible classes.

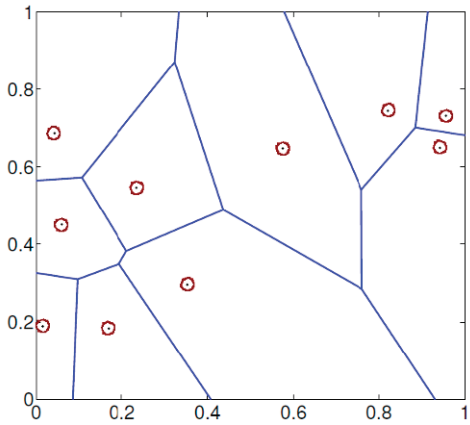

Figure 1. Abstract representation of a simple conceptual space with 2 dimensions which is populated by 10 concepts (circled points). Through generation of a Voronoi diagram the boundaries of the concepts can be illustrated. Within CS these boundaries are not explicitly defined or represented, but illustrate the assignment of presented examples as belonging to a single concept through the application of a distance measure (Equations 1 and 2).

2.1.1. Mechanisms

The notion of prototypes comes naturally to conceptual space modelling, as the distance metric functions as a metric of typicality. Distance d_{xy} between a prototype x and an example y takes the general form:

$$d_{xy} = \left(\sum_{i=1}^{N} w_i |x_i - y_i|^r \right)^{\frac{1}{r}} \quad (1)$$

where r denotes the type of metric with $r = 2$ for the Euclidean distance (or $r = 1$ for Manhattan distance) and $w = 1.0$ the weight of the dimension (with this parameter value, all dimensions are treated equally, which is a reasonable choice given no *a priori* domain knowledge). To do justice to psychological evidence of how people tend to rate concepts [21, 25], the distance is converted into a similarity measurement:

$$s_{ij} = e^{-c d_{ij}} \quad (2)$$

where the similarity s between concept prototypes i and j is computed as an exponentially decaying function of distance, where $c = 1.0$ is a sensitivity parameter.

2.2. Distributed Associative Interactive Memory (DAIM) Model

The DAIM model operates on a set of four functional principles derived from the operation of memory within biological systems, embedded within the context of a wider cognitive system [5, 32]. These principles are [32]: (1) memory as being fundamentally associative; (2) memory,

rather than being a passive storage device, is an active component in cognition through activation dynamics; (3) memory as having a distributed structure; and finally (4) activation-based priming as subserved by the first three points. The DAIM model has been implemented so as to embody each of these principles in a computational architecture (Figure 2).

Assuming that this system is embedded within a wider agent cognitive system with multiple sensory and motor *modalities*, associations may be seen to form based on the experiences of the agent between *units* of processing in these modalities (i.e. a localist representation scheme) in a Hebbian manner, which subsequently form the substrate for activation dynamics. Prior experience as encoded in associative networks, i.e. memory, thus plays an active role in the generation of ongoing behaviour through the mechanism of *priming*, which is the reactivation of modality-specific localist representations on the basis of existing associations. These principles (or variations thereon) have been used to provide candidate mechanisms for a wide range of cognitive phenomena, from visual recognition and analogies [1, 15], to episodic memory, language development and social interaction [5]. The application of these principles to conceptual has recently received support from neuroscientific studies which emphasise the distributed nature of conceptual representations [14, 18], thus motivating the present investigation.

2.2.1. Mechanisms

The computational implementation of the DAIM model is based on an extension to an Interactive Activation and Competition (IAC) model of face learning [7]. An explicit encoding for associations is thus used: an association is represented by an object (in the context of Object-Oriented Programming), following the approach taken in [2]. This model differs from standard IAC models (such as [17]), and their learning extensions (e.g. [7]) in four main respects. Firstly, rather than committing to defining a hub of connectivity, DAIM allows all pools of property units (i.e. modalities) to link to other units in any other modality: i.e. point-to-point connectivity (Figure 2). Secondly, weights are updated incrementally at run time, rather than as a batch process only when certain activation stability criteria are met (e.g. [7]). Thirdly, there is the capacity to create new associative links at run-time, rather than only enabling the adaptation of a structure initialised *a priori*. Finally, in contrast to standard IAC implementations, in the DAIM implementation used for this study, mutual inhibition between the units of a modality are not implemented (although the capacity for this functionality is present).

There are two main mechanisms present in DAIM: activation spread, and weight update. Activation is taken to be a scalar in the range $[a_{min}, a_{max}]$ where $a_{min} = -0.2$ and $a_{max} = 1.0$. A resting activation level is defined, which is the steady-state activation level of a unit in the absence of stimulation: $a_{rest} = -0.1$. Similarly, weights (of associative links) are taken to be scalars in the range $[w_{min}, w_{max}]$ where $w_{min} = -1.0$ and $w_{max} = 1.0$; the initial weight of associative links upon creation is defined as $w_{init} = 0.2$. A new associative link is created between two *units* in different *modalities*, *iff* they are the most active units in their respective modalities, and such a link does not already exist (see Figure 2). The net activation input to each unit is derived as follows, where ext_i is the activation derived from an 'external' source (input to a modality); $\xi_g = 0.6$ is a parameter controlling the proportion of externally derived activation used; int_i is the activation derived from within DAIM (activation from other modality units); and $\zeta_g = 0.3$ controls the influence of int_i:

$$net_i = \left(\xi_g \times ext_i\right) + \left(\zeta_g \times int_i\right) \tag{3}$$

This is based on the derivation of the activation spread from within the DAIM system (int_i), which is calculated as follows (on every time-step),

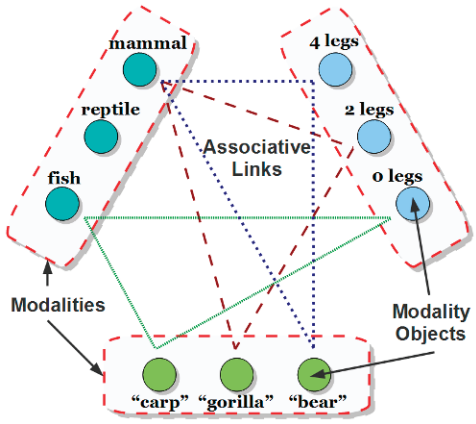

Figure 2. The structure of DAIM: shown is a subset of the structures acquired during the training process. Associative links are formed between the objects of modalities as these conjunctions are experienced. For example, the animal with label "gorilla" has been encoded as a conjunction of *mammal* and *2 legs* with this label.

where w_{ij} is the weight of an associative link linking unit i to unit j in another modality, and out_j is the activation of the linked unit j:

$$int_i = \sum_j w_{ij} \times out_j \tag{4}$$

The calculated net_i effectively encodes the net effect of all inputs to each individual unit, on each time-step, with the result being 'excitatory' if $net_i > 0$, and 'inhibitory' if $net_i < 0$. On this basis, the change in activation level of each unit may be updated (Δa_i), where $\delta_g = 0.1$ determines the proportional decay in activation level:

$$\begin{aligned} If(net_i > 0) : \Delta a_i &= net_i\,(a_{max} - a_i) - \delta_g\,(a_i - a_{rest}) \\ else : \Delta a_i &= net_i\,(a_i - a_{min}) - \delta_g\,(a_i - a_{rest}) \end{aligned} \tag{5}$$

The bounded weight update mechanism is based on that derived by Burton et al [7], with weight update magnitude being driven by the activation magnitudes of the linked units, where $\lambda_g = 0.01$ is the learning rate:

$$\begin{aligned} If(a_i a_j > 0.0) : &\quad \Delta w_{ij} = \lambda_g a_i a_j \left(1 - w_{ij}\right) \\ else : &\quad \Delta w_{ij} = \lambda_g a_i a_j \left(1 + w_{ij}\right) \end{aligned} \tag{6}$$

This weight update mechanism operates under the additional constraint (not included in the original Burton et al formulation) that:

$$if\ \left((a_i < 0) \cap (a_j < 0)\right) : \Delta w_{ij} = 0 \tag{7}$$

This counteracts the effect of the negative activation resting value (a_{rest}) gradually increasing all weights, and allows the learning mechanism to be always switched on and incremental at run-time, in contrast to the original learning IAC formulation where weight updates are batch processed when certain activation stability criteria are fulfilled (e.g. [7]).

2.2.2. What makes DAIM a Developmental Architecture?

As described above, the DAIM model essentially provides an active associative substrate upon which activation dynamics operate. This substrate is subject to adaptation (e.g. associative link weight update) over the course of the interaction of the system as a whole with its environment: as such, adaptation and activation dynamics are inherently inter-dependent. It is thus reasonable to ask whether DAIM can be classed as being a developmental system, rather than 'merely' a learning system. We contend that it *is* subject to a developmental trajectory, since a fundamental feature of operation is the creation of the associative substrate itself based on experience (through the creation of new associative links), in addition to its subsequent adaptation, which may be regarded as learning [4].

3. Modelling Conceptual Prototypes

To examine the acquisition of conceptual prototypes by the DAIM model, a dataset is applied and its classification accuracy with respect to that of the CS model is compared, by means of a 10-fold cross-validation procedure. In this context however, such a comparison only provides limited insight; what is also important is the extent and type of mis-classifications made. To further explore this issue, an additional analysis is conducted, specifically examining the prototype-like behaviour of the DAIM system with respect to the CS benchmark. The dataset chosen consists of a set of zoo animals. This dataset was chosen for this study due to its general familiarity: animals are readily classified by humans into groups of classes (mammal, bird, etc), with various animals viewed as more or less typical of a given class. Reflecting the functionalities of concepts described above (namely classification and typicality), this therefore enables the results to be intuitively assessed as well as quantitatively examined.

3.1. The Zoo Dataset

The data set that is the subject of this study is the zoo animal data set from the UCI Machine Learning Repository [9], which is a database of 100 named animals, each comprised of 17 properties. The majority of these properties are binary, such as 'has hair', 'is aquatic', or 'lays eggs'. The other two properties are categorical (animal class, which takes one of seven values: Mammal, Bird, Reptile, Fish, Amphibian, Insect, or Invertebrate), and scalar (number of legs; 0, 2, 4, 5, 6, 8). It can be seen that the distribution of instances over the classes is very uneven (Table 1), with overlapping classes leading to potential difficulties in classification [3]. However, one advantage of using a dataset such as this is that it enables an intuitive assessment of the system behaviours, in addition to the quantitative results that are obtained.

3.2. Experimental Procedure

The study reported in this paper has two distinct aspects. In the first (*Concept Identification*), the classification accuracy of the DAIM model is assessed using a 10-fold cross-validation scheme, and compared with the performance of the CS model. The emphasis of this analysis of classification is not just to assess the efficacy of classification, but also to assess, in relation to the CS model, where and why errors occur: this is necessary given the goal of accounting for human classification competencies with DAIM. The second aspect (*Prototype Functionality*) involves a deeper inspection of the manner in which classifications are made, specifically in relation to the prototype theory of human concept competencies. To this end, an additional split of the dataset is made, with the intention of comparing the classification performance of specific animal instances (Table 1). A detailed description of the procedures used for these two aspects may be found below (Section 3.2.3). The training and testing[1] procedures used for both parts of the study are the same, and so are described first for both the DAIM and CS models. Given the modelling nature of this study, the precise parameters used are of reduced significance compared to the mechanisms and functionality under test, where parameters are chosen through empirical selection: this means of parameter selection is consistent with that used in established IAC formulations, e.g. [7, 17].

Table 1. Class distribution in the complete zoo animal dataset, the 10-fold cross-validation results for CS and DAIM broken down by class (classification accuracy shown, two central columns); and the distribution of training and probe data used in the prototype functionality case study (two right-hand columns).

Class	Dataset number	CS success rate	DAIM success rate	Prototype Training (number)	Prototype Probe (number)
Mammal	41	1.00	1.00	39	2
Bird	20	1.00	1.00	16	4
Fish	13	1.00	1.00	12	1
Invertebrate	10	0.70	0.95	10	0
Insect	8	1.00	0.69	8	0
Reptile	5	0.60	0.70	4	1
Amphibian	3	1.00	0.33	3	0
Overall	100	0.95	0.94	92	8

3.2.1. Procedure for CS

The training of the CS model involves presenting the properties of all training instances to the system sequentially: there is no order effect. A conceptual space is set up based on the animal 'class' property such that each presented animal instance is projected to a point in this space so that a distance (and hence typicality) measurement can be determined between the instances and prototypes. It should be noted that the property 'number of legs' is normalised prior to training, to maintain equivalent ranges across all properties. Learning in the CS model is thus supervised, as the subject of the classification (animal class) is provided with the instance to be learned. These explicit prototype representations enable the classification of novel stimuli: during the probe phase, the probe instance properties are presented to the CS system, projected to the animal class conceptual space, and similarity measures to the prototypes present derived (see Equation 2).

3.2.2. Procedure for DAIM

For the DAIM model in this study, each of the animal instance properties constitutes a *modality*, with the features of each property (i.e. true/false, name, number of legs, animal class) constituting the modality *units* (please refer to Section 2.2 for the relation to the theory). Training in this case is thus unsupervised as all properties are treated equally.

[1] The term 'probe phase' is used to denote the testing part, following its use in psychology studies to indicate the part of the experiment where knowledge acquired over the course of the study is assessed.

The training procedure takes into account the temporal and iterative nature of the learning mechanisms (specifically weight update). Training takes the form of a sequence of instances to learn, where the order of the sequence may be seen as an analogue to the differing experience of multiple agents within the same environment (see Figure 3): the properties of an instance are presented to the system for 5 time-steps (during which time associative links are created and updated), followed by a period of 20 time-steps in which there is no input to the system (to ensure all activation decays before the next instance presentation). This presentation occurs as follows: an activation value of 1.0 is applied to all of the modality units corresponding to the properties of the animal instance; all other units receive no activation (i.e. 0.0). For the probe trials, the properties for each of the probe instances are presented to the system for 10 time-steps, followed by 60 time-steps of no input (Figure 3). The length of the probe stimulation is sufficient for the activations on the animal class properties to reach a steady state (see Figure 4). These steady state activation levels are then normalised and the resultant values are used as the basis of the results reported below.

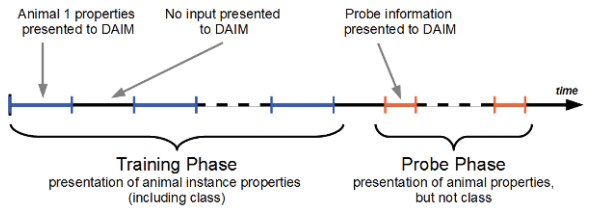

Figure 3. Training and testing DAIM: a sequence of animal instances are presented to DAIM (*Training* phase), separated by periods of no input. During the *Probe* phase, the properties of the animal instance are presented to DAIM; the activations on the class modality are read out (e.g. see figure 4). DAIM is trained multiple times with different animal instance orderings in the Training phase; see text for details.

Given that the learning mechanism in DAIM is incremental (as a result of the associative link creation and weight update mechanisms), the order of data presentation during learning influences the learned information, and hence the behaviour of the system. To assess the effects of varying presentation orders on the ability of the system to correctly classify probe instances, in each case the order of the training set was randomised to form ten separate training sets. Each of the ten resulting trained versions of the DAIM model may thus be regarded as having a different experience in the environment. In the probe phase the weight update mechanism was disabled to ensure that comparisons could be made across the probe animal instances on a common basis. For both aspects of the results, we now describe what the training and probe sets are constituted of.

3.2.3. Two Aspects of Investigation

The first aspect of this study (*Concept Identification*) looked at the capacity to identify concepts of CS and DAIM, by means of a 10-fold cross validation procedure. The zoo dataset was randomly partitioned, with the same partitions being used for both DAIM and CS. Furthermore, the order of the training set of DAIM for each fold was further randomised as described above. Therefore, a total of one hundred DAIM simulations were run, with results derived from each of the trained DAIM instances. A winner-takes-all mechanism was used on the produced activation values (based on magnitude, see e.g. Figure 4) to determine the classification result: see Section 4.1.

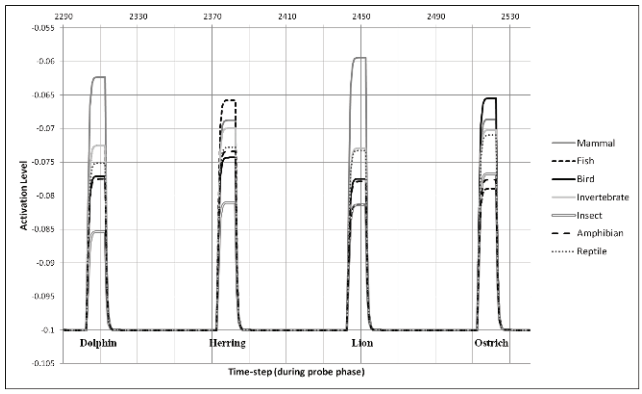

Figure 4. Example activation profiles for the 'Animal Class' property during the probe phase of one DAIM run, showing four probes (the properties of Dolphin, Herring, Lion and Ostrich). 'Resting' activation level is -0.1; rising activity is due to activation spread on the substrate of created associations between properties. Steady-state activation levels remain constant for at least 3 time-steps. These activation levels are normalised for the results.

The second aspect of this study (*Prototype Functionality*) is a closer examination of the prototype-like functionality exhibited by DAIM, using CS as a benchmark. For this, both models were trained on a further and separate partition of the zoo dataset. Eight animal instances were set aside for probe trials: lion, dolphin, seasnake, herring, ostrich, parakeet, penguin and pheasant. The remaining 92 instances were used as training data. For DAIM, this involved generating 10 randomly ordered training sets. The distribution of animal classes used in the training and probe sets for this procedure are shown in Table 1. The choice of instances to use as probes was based on the utility in illustrating the functionality of similarity to prototype judgements that are required for the typicality property of human concept competencies: this is discussed in Section 4.2.

4. Results

To elucidate the performance of DAIM with respect to CS, we first examine the classification performance (Section 4.1). Whilst the cross-validation procedure provides an overall perspective, we also examine how the DAIM activation levels allow similarity ratings between different classes to be made, by looking at classification performance of a sub-set of the zoo dataset. We then more closely examine the ability of DAIM to make typicality judgements between multiple instances of the same (correctly classified) class (Section 4.2), using birds as a case study.

4.1. Concept Identification

The first assessment that is made is the classification accuracy of the CS and DAIM models. The results show a high overall classification rate (Table 1): 95% for CS, and 94% for DAIM. The breakdown of accuracy by class is also instructive, showing that the classes for which there are many examples (e.g. mammals and birds, both with 100%) have a higher rate of successful classification than those with fewer examples (e.g. reptile and amphibian).

Table 2. Animal instances where CS and DAIM made classification errors, and how the errors are made: compare with Table 1 for overall correct classification rate. For DAIM, results of 10-fold cross-validation shown.

Animal Class	Number in class	CS errors	DAIM errors
Mammal	41	no errors	no errors
Bird	20	no errors	no errors
Fish	13	no errors	no errors
Invertebrate	10	two animals classified as *insects*, one as *reptile*	scorpion classified as both *invertebrate* and *reptile*
Insect	8	no errors	confused for *invertebrate* in 32% of cases
Reptile	5	seasnake classified as *fish*, and tortoise as *bird*	seasnake classified as *fish*
Amphibian	3	no errors	confused for *reptile* in 66% of cases

An examination of how the mis-classifications are made is also instructive (Table 2). For example, for DAIM, classification accuracy of amphibians is low, with only one from three instances being correct. However, DAIM consistently identified the other two amphibians as reptiles, which is a closer characterisation of amphibians than, say, mammals or birds. Similarly, the errors made for insects were due to classification as invertebrates. Finally, of the five reptiles present in the dataset, only one was consistently classified as a fish: this is the case of the seasnake, which is further explored below (Section 5). Given the uneven distribution of animals across the classes, it may be seen that (as is naturally expected) the mis-classifications occur more for those classes with fewer instances (Table 2). For CS, a similar pattern of errors occurs, with for example invertebrates classified as insects or reptiles, and seasnake classified as a fish. Taken together, these results support the notion that DAIM approximates the behaviour of CS in quality (the types of errors made) as well as quantity (overall classification rate).

Going further than examining where mis-classifications were made, we may examine how the identification of one animal class over another is achieved: in terms of concept prototype theory, this indicates how the learned prototypes relate to one another. For this question, we re-examine a subset of the zoo dataset (see right side of Table 1). For the CS model (Figure 5) the classifications of lion and herring are clear, in that the correct class has a far greater typicality rating than the other classes. Similarly, dolphin is correctly classified as a mammal, although the typicality rating for fish is comparable. Finally, seasnake is incorrectly classified, with fish and amphibian being identified to a similar extent instead.

For the DAIM model, the results show the mean normalised activation values across ten randomly ordered runs (see final paragraph of Section 3.2, and Figure 6). Firstly, it can be noted that there is a high level of consistency of results across the 10 training set orders (as evidenced by the small 95% CI's), indicating that while there is an order effect, it does not disrupt classification accuracy. In accordance with the CS results, lion, herring and dolphin are correctly classified. However, seasnake is not classified correctly, with similar activation values for mammal, fish and invertebrate. This is quantitatively the most divergent result in comparison with those derived from the CS model, although qualitatively, the fact that seasnake is misidentified, with a

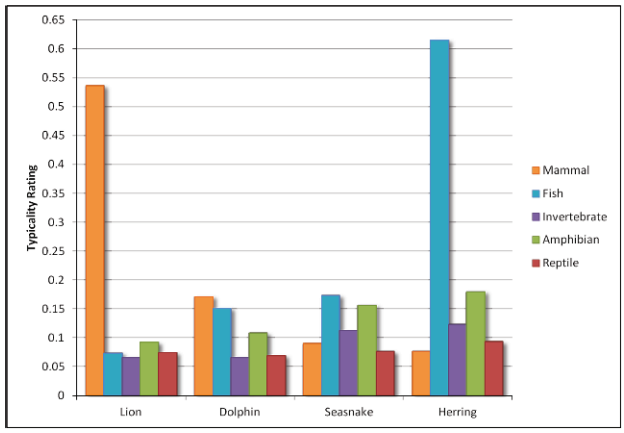

Figure 5. Conceptual Space model classification results for four animals. Five of the seven available animal type categories are shown, for the purpose of clarity. All animals are classified correctly, except seasnake (Reptile): this is classed as a fish, although the rating for amphibian is similarly high.

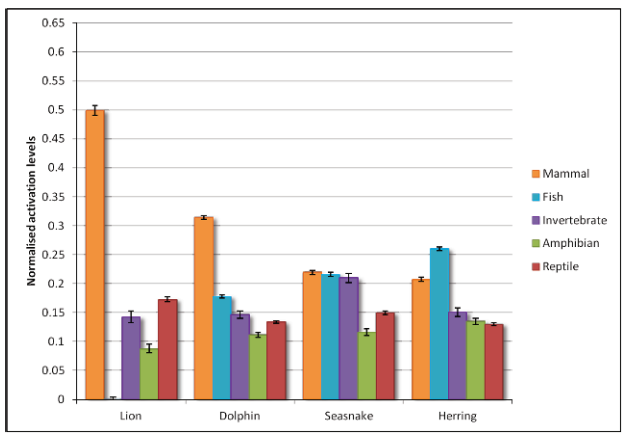

Figure 6. DAIM model classification results for four animals. The same five categories are used as in Figure 5. The results show the mean of ten randomly ordered datasets (see main text for details): the error bars show 95% confidence intervals. As for the CS model, all animals are classified correctly except seasnake.

number of possible (indeed intuitively plausible) alternatives present in both cases. Additionally, it can be noted that the relative magnitude of the activation of mammal for seasnake and herring is higher for DAIM than for the CS model.

4.2. Prototype Functionality

The second assessment that can be made is the degree to which the respective models are able to determine the typicality of novel input animal instances to a prototype. Assuming that classification is correctly performed, the question here is whether the two models can produce a measure of how close a presented animal instance is to a learned concept prototype. This measure of typicality is explicitly implemented in the CS model (cf. Eq. 1), but not for DAIM, where all information is

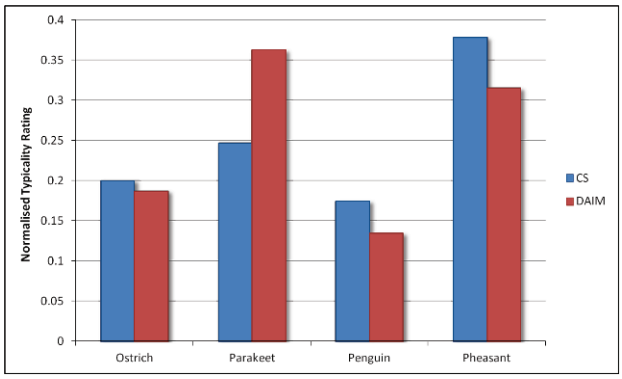

Figure 7. Comparing the relative typicality of four birds, for both DAIM and CS. All four were correctly classified as birds by both models; this shows that for both models parakeet and pheasant are more typical birds than penguin and ostrich.

maintained in a distributed state: any such assessment must be made on the basis of relative activation levels (e.g. Figure 4).

A comparison of four birds is conducted. Each bird (ostrich, parakeet, penguin and pheasant) is presented as one of the eight probes in the second simulation set of the study. For both the CS and DAIM models, each of these birds are classified correctly. For the CS model, typicality ratings are calculated, and normalised across the four values (Figure 7). For DAIM, the activation values of the bird class for each of the four instances are taken and normalised. It can be seen that in this case, parakeet is identified as the most typical bird, and that there is a clear difference for DAIM between the pair parakeet-pheasant (indicating that they are highly typical of the bird concept), and ostrich-penguin (being relatively atypical). A similar qualitative pattern and division of the two pairs can be seen in the CS results (Figure 7), even though pheasant is identified as the most typical bird instance.

5. Discussion

The presented results show that there is similarity between the behaviours of the CS and DAIM systems for classification of the novel animal instances, with classification rates of 95% and 94% respectively. Additionally, there is a qualitative similarity of the DAIM system performance in typicality ratings to those derived by the CS model, despite the fact that the means of calculating and assessing typicality and similarity fundamentally differ. In a comparison of the bird typicality ratings from the two models (Figure 7), while the actual order of ratings differs (with parakeet being most typical of a bird for DAIM, and pheasant for CS), the indication that these are more typical than either penguin or ostrich is clear (and thus matching intuition). There are a number of potential sources for the differences, including the exponential-based calculation of similarity for CS, and the order-dependent effects for DAIM. However, that such strong similarities exist for both classification and prototype-based similarity assessments provides support for the notion that the mechanisms that DAIM makes use of can account for these two fundamental features of concept functionality. In addition to these observations, a number of other issues merit further consideration.

5.1. The Case of the Seasnake: Errors and Prototypicality

The case of the mis-classified seasnake raises a number of questions. That both CS and DAIM fail to classify it correctly may be an indication that there is some inherent ambiguity resulting from the dataset itself. Indeed, if the distribution of animal classes is considered (Table 1), there are very few examples of reptiles in comparison to mammals for example. The manner in which mis-classifications were made for both CS and DAIM reflect to some degree the overlapping properties of the seasnake with animals from other classes, notably fish. The inclusion of mammal into this consideration for the DAIM results may reflect the larger proportion of mammals in the dataset on the incremental nature of the weight update mechanism (e.g. *if the majority of animals seen are mammals, then the chances are that in the absence of distinguishing features, a novel animal may also be a mammal*). Nevertheless, there is an indication in this behaviour that even with mistaken classifications, there is the possibility for the outcome (i.e. the distribution of activity over multiple animal classes) to be of utility in further processing, by, for example, providing a set of hypotheses that can be used as the basis for further disambiguation actions.

This effect is related to the prototype effect, where an assessment can be made not just with regards to whether a presented example is within a category or not, but also how close it is to any known category. The case study with the birds showed that DAIM can achieve this assessment in a manner consistent with CS but using a developmental, unsupervised account: all correctly identified as birds in the first instance, a distinction could be made between different birds regarding how typical they were of the concept. When errors are made, this same mechanism allows different categories to be identified as potential candidates (e.g. Figures 5 and 6). As mentioned above, it is this mechanism that has potentially profound implications for further cognitive processing (i.e. the consequences for cognitive architecture, Section 5.3).

5.2. Robustness to Varying Experience

The results of the classification task for the DAIM model (Table 1, Figure 6) indicate that the classification performance is robust to presentation order within the learning phase. The cross-validation demonstrates this by showing that DAIM exhibits a high degree of consistency of classification among the ten randomly ordered datasets for each fold (be it a correct or incorrect classification). The random orderings of the training data may be considered in this context as an analogue of the varying experiences of multiple agents in an environment with equivalent statistical properties. There were only two exceptions to this consistency: scorpion (classified as reptile in half of the cases, and invertebrate the other half), and termite (half insect, half invertebrate).

In the context of developmental systems, the importance of a trajectory based on experience is typically emphasised. This result supports the notion that even with unique experiences, there is the capacity for the statistical regularities of the environments shared by agents to lead to robust conceptual categories for those agents, in support of coordination through inter-agent interaction: indeed, such an extension to the present work is of interest (e.g. [12, 13]).

5.3. Concepts and Cognitive Architecture

The zoo dataset represents relatively abstract information, and thus somewhat removed from the basic sensory data that an autonomous robotic agent would necessarily deal with. However, the study presented in this paper nevertheless demonstrates how the functionality of prototype conceptual processing can be achieved from (albeit ab-

stracted) multi-modal data, and is therefore an informative illustration of the applicability of the principles of operation that DAIM embodies. There are two further reasons why the use of the zoo dataset was beneficial. Firstly, using a subject matter that is generally and intuitively familiar enables the results to be easily interpretable. Since we are concerned with the formation of concept prototypes in humans, this is a useful feature when considering the results, in a way that a far more abstract dataset would fail to achieve. Secondly, this dataset has been used in another human-centred study, described below, which provides a useful benchmark in the interpretation of these results in a broader context.

This same zoo dataset was used to teach a robot animal concepts by a human tutor through social interaction [13], in an investigation related to the relatively novel paradigm of socially-guided machine learning [30]. In this work, the robot benefits from active modulation of the interaction dynamics, enabling it to shape its own learning experience. Tutor and robot learner engaged in a series of Language Games [28], through which the learner gradually acquired animal concepts. Through the expression of social cues, the robot was able to influence the human tutor into providing a learning experience that was more effective, compared to a control condition in which the robot did not express any preferences through social cues. This illustrates how robot learning might be embedded within social interaction that is natural for people to engage in, demonstrating the integration of multiple cognitive competencies. The use of the zoo dataset provides learning material that is intuitive for human tutors; yet, from the standpoint of artificial learning, it is challenging enough to highlight the merits of social augmentation of the learning process.

The DAIM model is described above as being a system that has a developmental trajectory in terms of increasing competencies with increasing interactions with its environment. As part of a wider cognitive system, the principles upon which it operates enable DAIM to bias, influence and/or modulate ongoing system behaviours using the mechanism of priming [5]. In this study it is seen that allowing activation to persist in multiple units may be beneficial for processing in a wider cognitive context, particularly where a clear classification fails in the first instance. By demonstrating that DAIM can account for (at least some central aspects of) conceptual functionality in a sub-symbolic manner, there is a potential reduction in reliance on explicit symbolic constructs. This indicates that there may be a fundamental mechanistic integration of conceptual competencies and their development within wider cognitive processing (e.g. [19]), within an autonomously developmental framework. The human tutor/robot learner study cited above provides one demonstration of the necessity to consider the fundamental relationship between a particular cognitive competence, in this case conceptual processing, and wider cognitive and behavioural processing. This perspective on broadening the scope of cognition presents itself as a promising avenue for further investigation.

6. Conclusion

The purpose of this study was to explore whether a developmental memory system (DAIM) informed by neuropsychological principles could account for aspects of human-like conceptual processing. Using a CS model as a benchmark of human performance, the results obtained indicate that the DAIM model can reproduce two fundamental properties of concepts: categorisation and prototype-based similarity assessments. It does so using a distributed representation scheme operating on the principles of association and activation dynamics, which are consistent with, and have been used to account for, a wide range of other cognitive competencies. This allows conceptual competencies to be viewed in the context of wider cognitive processing and in the framework of development, a perspective of particular relevance to the creation of autonomous robotic agents. While there remain a number of open questions, this study has provided evidence in support of the embedding of conceptual competencies within the developmental memory-centred perspective on cognition.

Acknowledgment

This work is funded by the EU FP7 ALIZ-E project (grant 248116).

References

[1] M. Bar, *The proactive brain: using analogies and associations to generate predictions*, Trends in cognitive sciences, vol. 11, no. 7, pp. 280–289, (2007)

[2] P. Baxter, *Foundations of a constructivist memory-based approach to cognitive robotics*, PhD. Thesis, University of Reading, U.K., (2010)

[3] P. Baxter, J. de Greeff, R. Wood, T. Belpaeme, *"And what is a Seasnake?": Modelling the Acquisition of Concept Prototypes in a Developmental Framework*, 2nd joint International Conference on Developmental Learning (ICDL) & Epigenetic Robotics, San Diego, USA, IEEE Press, (2012)

[4] P. Baxter and W. Browne, *Memory as the substrate of cognition: a developmental cognitive robotics perspective*, 10th International Conference on Epigenetic Robotics, pp. 19–26, (2010)

[5] P. Baxter, R. Wood, A. Morse, and T. Belpaeme, *Memory-Centred Architectures: Perspectives on Human-level Cognitive Competencies*, AAAI Fall 2011 Symposium on Cognitive Systems, pp. 26–33, (2011)

[6] H. Branigan, M. Pickering, J. Pearson and J. McLean, *Linguistic alignment between people and computers*, Journal of Pragmatics, vol. 42, no. 9, pp 2355–2368, (2010)

[7] A. Burton, *Learning new faces in an interactive activation and competition model*, Visual Cognition, vol. 1, no. 2, pp. 313–348, (1994)

[8] A. Chella, S. Gaglio, and R. Pirrone, *Conceptual representations of actions for autonomous robots*, Robotics and Autonomous Systems, vol. 34, no. 4, pp. 251–263, doi:10.1016/S0921-8890(00)00121-4, (2001)

[9] A. Frank and A. Asuncion, *UCI Machine Learning Repository*, http://archive.ics.uci.edu/ml (*accessed 15/12/2012*), Irvine, CA: University of California, School of Information and Computer Science, (2010)

[10] P. Gärdenfors, *Conceptual Spaces: The Geometry of Thought*, Cambridge, MA: MIT Press, (2000)

[11] P. Gärdenfors, and M. Warglien, *Using Conceptual Spaces to Model Actions and Events*, Journal of Semantics, doi:10.1093/jos/ffs007, (2012)

[12] J. de Greeff, F. Delaunay, and T. Belpaeme, *Human-Robot Interaction in Concept Acquisition: a computational model*, International Conference on Development and Learning, pp. 1–6, (2009)

[13] J. de Greeff, F. Delaunay, and T. Belpaeme, *Active robot learning with human tutelage*, 2nd joint International Conference on Developmental Learning (ICDL) & Epigenetic Robotics, San Diego, USA, IEEE Press, (2012)

[14] M. Kiefer, and F. Pulvermuller, *Conceptual representations in mind and brain: theoretical developments, current evidence and future directions*, Cortex, vol. 48, no. 7, pp. 805–25. doi:10.1016/j.cortex.2011.04.006, (2012)
[15] R. Leech, D. Mareschal, and R. Cooper, *Analogy as relational priming: a developmental and computational perspective on the origins of a complex cognitive skill*, Behavioral and Brain Sciences, vol. 31, no. 4, pp. 357–78; discussion 378–414, (2008)
[16] E. Margolis and S. Laurence, *Concepts: Core Readings*, MIT Press, (1999)
[17] J. McClelland and D. Rumelhart, *An Interactive Activation Model of Context Effects in Letter Perception: Part 1, an account of basic findings*, Psychological Review, vol. 88, no. 5, pp. 375–407, (1981)
[18] C. McNorgan, J. Reid, and K. McRae, *Integrating conceptual knowledge within and across representational modalities*, Cognition, doi:10.1016/j.cognition.2010.10.017, (2010)
[19] A. Morse, J. De Greeff, T. Belpaeme, and A. Cangelosi, *Epigenetic Robotics Architecture (ERA)*, IEEE Transactions on Autonomous Mental Development, vol. 2, no. 4, pp. 325–339, (2010)
[20] G. Murphy, *The Big Book of Concepts*, MIT Press, (2002)
[21] R. Nosofsky, *Attention, similarity, and the identification/categorization relationship*, Journal of Experimental Psychology-General, vol. 115, no. 1, pp. 39–57, (1986)
[22] M. Pickering, and S. Garrod, *Toward a mechanistic psychology of dialogue*, The Behavioral and brain sciences, vol. 27, no. 2, pp 169–190, (2004)
[23] R. Pfeifer, M. Lungarella, and F. Iida, *Self-organization, embodiment, and biologically inspired robotics*, Science, vol. 318, no. 5853, pp 1088–1093, (2007)
[24] E. Rosch, *Natural categories*, Cognitive Psychology, vol. 4, no. 3, pp. 328–350, (1973)
[25] R. Shepard, *Toward a universal law of generalization for psychological science*, Science, vol. 237, no. 4820, pp. 1317–1323, (1987)
[26] F. Shic, and B. Scassellati, *Pitfalls in the Modeling of Developmental Systems*, International Journal of Humanoid Robotics, vol. 4, no. 2, pp. 435–454, doi:10.1142/S0219843607001084, (2007)
[27] E. Smith and D. Medin, *Categories and Concepts*, vol. 4, Harvard University Press, (1981)
[28] L. Steels, *The Talking Heads Experiment. Volume 1: Words and Meanings*, Laboratorium, Antwerpen, (1999)
[29] R. Sun, *Desiderata for cognitive architectures*, Philosophical Psychology, vol. 17, no. 3, pp 341–373, (2004)
[30] A. Thomaz, *Socially Guided Machine Learning*, PhD thesis, MIT, (2006)
[31] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen, *Autonomous mental development by robots and animals*, Science, vol. 291, pp. 599–600, (2001)
[32] R. Wood, P. Baxter, and T. Belpaeme, *A review of long-term memory in natural and synthetic systems*, Adaptive Behavior, vol. 20, no. 2, pp. 81–103, (2012)