Proceedings of the Post-Graduate Conference on Robotics and Development of Cognition
10-12 September 2012, Lausanne, Switzerland

From Penguins to Parakeets: a Developmental Approach to Modelling Conceptual Prototypes

Joachim de Greeff, Paul Baxter, Rachel Wood and Tony Belpaeme

Centre for Robotics and Neural Systems
University of Plymouth, United Kingdom
joachim.degreeff@plymouth.ac.uk

Abstract

The use of concepts is a fundamental capacity underlying complex, human-level cognition. A number of theories have explored the means of concept representation and their links to lower-level features, with one notable example being the Conceptual Spaces theory. While these provide an account for such essential functional processes as prototypes and typicality, it is not entirely clear how these aspects of human cognition can arise in a system undergoing continuous development - postulated to be a necessity from the developmental systems perspective. This paper seeks to establish the foundation of an approach to this question by showing that a distributed, associative and continuous development mechanism, founded on principles of biological memory, can achieve classification performance comparable to the Conceptual Spaces model. We show how qualitatively similar prototypes are formed by both systems when exposed to the same dataset, which illustrates how both models can account for the development of conceptual primitives.

Index Terms: Concepts, prototypes, typicality, Conceptual Spaces, Distributed Associative and Interactive Memory

1. Introduction

For a cognitive system to be able to perform at a level that is comparable to humans, it should be able to form conceptual structures as part of its knowledge representation capacities. As concepts are recognised as being important for many aspects of cognition, it is paramount for an artificial system to be able to model conceptual knowledge, including the formation of prototypes.

In this paper we examine two frameworks for modelling human knowledge; one is based on Conceptual Spaces (CS) [1] and the other, Distributed Associative and Interactive Memory (DAIM), is centred around the distributed nature of human memory and the temporal aspects of its functioning [2, 3]. As they are focussed on different aspects of human knowledge these frameworks have both virtues and drawbacks. CS inherently models knowledge as summary representations which makes it natural to model some of the more generic properties of concepts. However, a CS is a rather static structure and from a developmental perspective it is less clear how well a CS would capture conceptual learning over time. Also, there are no inherent temporal aspects in the model that could account for some of the temporal aspects of human memory, thus a conceptual space is more abstract as a model of human cognition. DAIM on the other hand takes a more developmental approach and emphasises the low level associative and temporal properties of human knowledge acquisition. The question of reconciliation of the two approaches thus arises: can the developmental DAIM perspective be used to account for the structures and functions hypothesised by CS models? This paper seeks to address this question by applying both approaches to the same data set, to assess the compatibility of DAIM with CS.

As an example case, we explore the ability of both frameworks to model an aspect that is considered fundamental to human-like knowledge representation, namely the formation of prototypes which display *typicality* [4]. The observation by Rosch that many everyday concepts are prototypical in nature challenged the established notion in cognitive science that concepts could be modelled using logical definitions[1]. Rosch showed that many concepts cannot be logically defined because they show typicality, that is, people judge certain instances of a specific concept to be more typical than others. For example, for the concept BIRD, a robin is thought to be more "bird-like" than a penguin, a banana is more typical for FRUIT than a pomegranate etc. It turned out that instances of a concept exhibit a graded membership to an idealised prototype, so that some instances are seen as more typical of the concept than others.

Theories advocating this prototypical view of concepts have been around for quite a while with many different flavours [5, 6], but the general gist is that concepts are represented as some kind of idealised version of the specific concept. So, for the concept BIRD people would have an idea of the idealised bird, and match any encounters they have in the real world to this prototype version. The more similar a particular observation is to the prototype, the more they are inclined to assign this observation as belonging to the prototype. It may seem unlikely that all members of BIRD could be represented by one single prototype, given the wide variety of birds. So a prototype should be thought of as a *summary representation*, which specifies the properties of the concept, where some properties are more important than others. These properties are not strictly necessary, but rather they describe what members of the concept tend to have. The process of identifying an object in the world entails a matching to known prototypes. This matching takes the form of a similarity measurement, rather than a logical "does it ticks the boxes?" type of analysis. A prototypical account provides a more naturalistic explanation of human data than a definitional approach.

[1]The idea that concepts can be represented as a list of logical definitions which specify necessary and sufficient conditions is commonly known as the Classical Theory.

2. Theory of the two frameworks

2.1. Theory of Conceptual Spaces

A conceptual space consists of a geometrical representation in vector space along various quality dimensions. A CS is a collection of one or more domains (like colour, shape, or tone), where a domain is postulated as a collection of inseparable sensory-based quality dimensions with a metric. Examples of quality dimensions are weight, temperature, brightness, pitch, loudness, and RGB values. For instance, to express a point in the colour domain using RGB encoding, the different quality dimensions *red*, *green*, and *blue* are all necessary to express a certain colour and are therefore inseparable. Other domains may consist of one or more quality dimensions. 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, an instance of the concept RED may be represented as a point (255, 0, 0) in the RGB colour domain.

Crucially 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 a lot of dimensions the Euclidean distance may be the most appropriate one, but the Manhattan distance can also be used.

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

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

where r denotes the type of metric with $r = 1$ for the Manhattan distance and $r = 2$ for the Euclidean distance and w an optional weight of the dimension. To do justice to psychological evidence of how people tend to rate concepts [7, 8], we can convert the distance 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.

Within a conceptual space we can model the learning of prototypes by exposing the model to examples with associated labels. After the learning the model is able to classify new examples as belonging to some known class, and specify how typical the example is, i.e. to what extent it belongs to the class and to other learned classes.

2.2. Theory of the Distributed Memory Model

The DAIM system operates on a set of functional principles derived from the operation of memory within biological system, embedded within the context of a wider cognitive system [9, 3]. These are as follows [3]: (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. A DAIM model has been implemented that embodies each of these principles of operation.

Assuming that this memory system is embedded within a wider agent cognitive system with multiple sensory and motor modalities, associations may be formed based on the experiences of the agent, which subsequently form the substrate for activation dynamics. Prior experience as encoded in associative networks, i.e. memory, thus play an active role in the generation of ongoing behaviour through the mechanism of priming, which is the reactivation of modality-specific representations on the basis of existing associations. These principles may be used to provide candidate mechansims for a wide range of cognitive phenomena, from visual recognition and analogies [10, 11], to episodic memory, language development and social interaction [9].

In this study, the notional 'embodiment' of the DAIM system is modelled by an idealised set of inputs i.e. the properties given in the dataset. Associations are formed between input properties, on the basis of activation dynamics (where a high activation level is assigned to a property that is present). These associations have a weight value that is manipulated throughout the operation of the system. This introduces a significant temporal effect, in that an association is continually subject to change based on the relative activation levels of the things it associates, using a Hebbian-like update mechanism. Thus, by extension, the order of learning also has effect on the behaviour of the system.

Implementation of the model is based on an extension to an Interactive Activation and Competition (IAC) model of face learning [12], and uses an explicit representation for associations: i.e. an association is encoded as an object[2], following [13]. While details of this implementation are excluded here due to space constraints, the following description outlines the primary mechanisms.

The weight update mechanism incorporates both Hebbian and anti-Hebbian rules, and essentially has the effect of turning the DAIM implementation into a pseudo-correlation engine, in which the strength of the weights encoding conjunctions of input features essentially reflects the correlation of those features based on prior experience. It should be noted that this is not a correlation in the proper sense, but only an analogue thereof, given the incremental update nature of the weight adjustment. Activation dynamics are also at play, with all input properties having an associated activation level. Activation for a particular property rises if it is present, and falls in the absence of stimulation (i.e. activation decay, to a negative activation 'resting' state). It should be noted that such stimulation can be sourced either from external stimulation, or from the result of activation flowing through already existing associations. A new association is formed between two properties if an association does not already exist, and if the activation of both properties is above zero.

3. Modelling prototypes using the dataset

To examine how both models are able to build conceptual structures that exhibit prototypes and typicality effects, we use the Zoo Data Set from the UCI Machine Learning Repository [14] which is a simple database containing 101 example animals with 16 different properties (like airborne, aquatic, predator etc.) divided into 7 classes. All properties are binary, except for the 'number of legs'. This property is normalized as to make it more in line with the other properties. Both models are exposed to a subset of this data (50 animals), and the resulting knowledge structures are compared by using a further non-

[2] In the context of Object-Oriented Programming.

Table 1: Typicality ratings of the CS model for the 10 examples from the test set.

example	MA	BI	FI	AM	INS	INV
moth	0.08	0.12	0.05	0.09	**0.49**	0.12
newt	0.11	0.14	0.21	**0.37**	0.09	0.13
octopus	0.06	0.07	0.09	0.13	0.12	**0.32**
opossum	**0.37**	0.08	0.08	0.11	0.07	0.06
oryx	**0.53**	0.07	0.06	0.07	0.07	0.05
ostrich	0.10	**0.25**	0.08	0.09	0.10	0.08
parakeet	0.07	**0.39**	0.07	0.07	0.13	0.06
penguin	0.08	**0.20**	0.11	0.13	0.07	0.10
pheasant	0.07	**0.57**	0.08	0.09	0.15	0.08
pike	0.08	0.08	**0.40**	0.13	0.05	0.10

overlapping subset of the data as probes[3]. For the CS the training data is provided with an associated word label that specifies the class, while for the DAIM system the class label and the class type as a numerical value are supplied in the same fashion as the 16 other properties, thus this system is exposed to 18 properties per example. The test data contains 50 examples, where the breakdown into classes is as follows: 24 MAMMAL, 7 FISH, 9 BIRD, 4 INVERTEBRATE, 1 AMPHIBIAN and 5 INSECT.

3.1. Assesment

After training the systems are tested with 10 examples that are not part of the training set. Based on the learned information, an assessment of which category a newly presented instance belongs to is made. To examine the typicality ratings for the different examples the similarity measure from equation 2 is used.

3.2. Conceptual spaces

Using a CS representation, for each item in the test set we obtain typicality ratings for all classes (see Table 1). All examples from the test set are classified correctly.

Focussing more on the BIRD class, we can clearly observe typicality effects, as shown in Figure 1. For the BIRD class, the pheasant is the most typical example, followed by the parakeet, the ostrich and finally the penguin. This is in line with human typicality ratings as for instance reported in [15], [16] and [17], except for the fact that pheasant is rated as more typical than parakeet. Upon closer inspection it turns out that the property ‘domestic’, which is true for a parakeet, is somewhat rare for BIRD and therefore the parakeet is rated as less typical. We speculate that the contrast with typicality ratings from human data is due to the fact that a property ‘domestic’ may not commonly be very prominent for people when classifying birds.

3.3. Distributed memory model

In order to assess the effects of presentation order, we run the DAIM system twice with the same dataset; once in alphabetical order of animal name, and the second in reverse alphabetical order. Because of the inherently temporal dynamics of the system, for this case study, the properties of each animal instance are presented for 5 time-steps[4] followed by a delay of 10 time-step in which no input is presented so that all activation can decay

[3]We chose a subset of the Zoo Data Set because to show the prototype effects the full dataset is not necessary. This is an arbitrary choice, we just choose the first 50 examples from a list in alphabetical order.

[4]A time-step resolution of 0.2s is used.

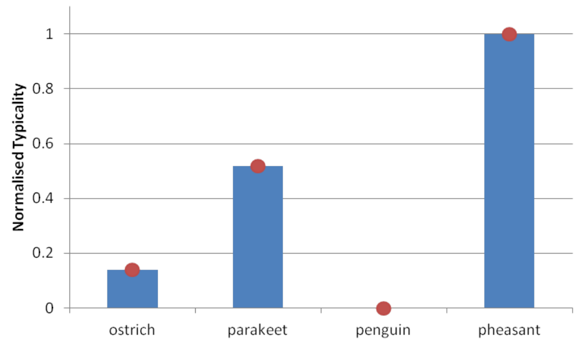

Figure 1: *Normalised typicality ratings of the CS model for the four probe trial birds for the BIRD class.*

to the resting state. For the probe trials, all of the properties for the unknown animal instances (except the name and type properties) are presented for 5 time-steps, with the activation levels on the type properties read out at the end of this period.

Table 2: Normalised results of the DAIM model for alphabetical presentation order: all correct.

PROBE	MA	BI	FI	AM	INS	INV
moth	0.01	0.08	0.00	0.22	**0.49**	0.19
newt	0.12	0.08	0.08	**0.52**	0.06	0.14
octopus	0.00	0.04	0.07	0.32	0.20	**0.34**
opossum	**0.63**	0.00	0.00	0.30	0.02	0.01
oryx	**0.78**	0.00	0.00	0.16	0.01	0.00
ostrich	0.01	**0.58**	0.00	0.30	0.08	0.01
parakeet	0.00	**0.55**	0.01	0.18	0.22	0.02
penguin	0.00	**0.50**	0.02	0.40	0.02	0.03
pheasant	0.00	**0.53**	0.01	0.25	0.15	0.02
pike	0.01	0.01	**0.42**	0.40	0.00	0.14

Table 3: Normalised results of the DAIM model for reverse-alphabetical presentation order: all but octopus are correct.

PROBE	MA	BI	FI	AM	INS	INV
moth	0.00	0.04	0.01	0.31	**0.41**	0.21
newt	0.05	0.03	0.25	**0.54**	0.00	0.13
octopus	0.00	0.04	0.13	**0.35**	0.09	0.34
opossum	**0.56**	0.00	0.02	0.36	0.01	0.02
oryx	**0.70**	0.00	0.02	0.23	0.01	0.00
ostrich	0.00	**0.52**	0.03	0.39	0.02	0.01
parakeet	0.00	**0.54**	0.04	0.25	0.08	0.03
penguin	0.00	**0.40**	0.10	0.39	0.03	0.04
pheasant	0.00	**0.48**	0.04	0.32	0.06	0.04
pike	0.01	0.02	**0.50**	0.35	0.00	0.09

The resulting typicality ratings, normalised, are shown in Table 2 and Table 3, for the two differently ordered data sets. Even though the typicality values differ for the two different data set orders, we can observe qualitatively similar results in terms of how the probe trials are classified. All but the octopus are assigned the same (correct) class and in this case of misclassification the typicality rating of the correct response is very close (0.35 and 0.34 respectively). Figure 2 shows typicality ratings for the BIRD class for the four bird examples in the probe trials. As can be seen, penguin and ostrich are hardly considered typical of BIRD, whereas parakeet and pheasant are rated as being much more typical. This is comparable to the result from CS, as displayed in Figure 1.

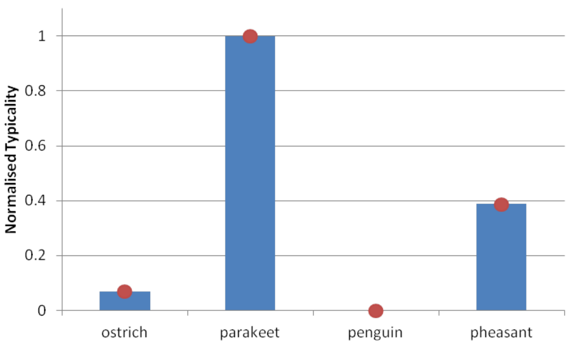

Figure 2: *Normalised mean typicality ratings of the DAIM model for the four birds presented in two probe trials. All four were classified correctly, but note that penguin and ostrich are far less typical of the bird concept than parakeet and pheasant.*

4. Discussion and conclusions

In this paper we have compared two knowledge representation frameworks for their ability to model conceptual prototypes. While Conceptual Spaces are quite suitable for this as they incorporate a notion of distance that can very easily be used as a typicality measure, it is less clear how this should happen in models that incorporate temporal aspects and are inherently distributed in nature. Whilst the temporal effects (as encountered in the order of presentation of the 50 instances to be learned) have a demonstrable effect on the behaviour of the system, the approach used in DAIM nevertheless demonstrates a robustness of ability to correctly classify the newly presented instances.

Furthermore, the DAIM results for the typicality ratings for the BIRD class exhibit prototypicality effects that are qualitatively similar to those obtained using a CS representation and to those found in human subjects. This shows the feasibility of the DAIM model, as the prototype effects are deemed important for conceptual modelling. Being inherently temporal and distributed, the use of memory models like DAIM can account for some of the more low-level functioning of the human memory, within a developmental framework (i.e. the history of interaction of the agent has a material effect on the competencies of the agent [2, 3, 13]). The comparison of the two memory conditions (normal and reverse order of data presentation) demonstrates that despite this sensitivity to interaction history (in this case order of presentation), there is nevertheless a robustness apparent in the outputs of the two trained systems. The fact that crucial aspect of modelling concepts, like prototypicality (which can more easily be modelled in a generic framework like CS) can be accounted for may be considered as an argument in favour of a distributed representation perspective; not being able to account for these aspects would constitute a shortcoming.

However, whilst the results of the DAIM system compare favourably with the standard CS implementation, it remains to be seen how such a distributed representation scheme can account for higher level concept manipulation. For example, the advantage of the CS representation scheme is the collapsing of multiple linked dimensions into a single point, that encodes a single concept or prototype. As such, it is readily available for further comparative operations with other concepts, and perhaps even higher-level processing. This property of the CS model is not so readily envisaged with the DAIM system given the entirely distributed nature of all acquired information.

Nevertheless, this study has demonstrated that some fundamental aspects of conceptual modelling can be accounted for in a distributed system that emphasises associative processes embedded within a complete cognitive system, engaged in ongoing development.

5. Acknowledgements

This work is funded by the CONCEPT project (EPSRC EP/G008353/1) and by the EU FP7 ALIZ-E project (grant 248116).

6. References

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

[2] P. Baxter, "Foundations of a constructivist memory- based approach to cognitive robotics," Ph.D. dissertation, University of Reading, U.K., 2010.

[3] 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.

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

[5] M. Posner and S. Keele, "On the genesis of abstract ideas," *Journal of experimental psychology*, vol. 77, no. 3p1, p. 353, 1968.

[6] S. Reed, "Pattern recognition and categorization," *Cognitive psychology*, vol. 3, no. 3, pp. 382–407, 1972.

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

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

[9] P. Baxter, R. Wood, A. Morse, and T. Belpaeme, *Memory-Centred Architectures: Perspectives on Human-level Cognitive Competencies*. AAAI Press, 2011, pp. 26–33.

[10] M. Bar, "The proactive brain: using analogies and associations to generate predictions." *Trends in cognitive sciences*, vol. 11, no. 7, pp. 280–9, Jul 2007.

[11] R. Leech, D. Mareschal, and R. P. Cooper, "Analogy as relational priming: a developmental and computational perspective on the origins of a complex cognitive skill." *The Behavioral and brain sciences*, vol. 31, no. 4, pp. 357–78; discussion 378–414, 2008.

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

[13] P. Baxter and W. Browne, *Memory as the substrate of cognition: a developmental cognitive robotics perspective*, 2010, pp. 19–26.

[14] A. Frank and A. Asuncion, "UCI machine learning repository," 2010. [Online]. Available: http://archive.ics.uci.edu/ml

[15] E. Rosch, "Cognitive representations of semantic categories." *Journal of Experimental Psychology: General*, vol. 104, no. 3, pp. 192–233, 1975.

[16] J. A. Hampton and M. M. Gardiner, "Measures of internal category structure: A correlational analysis of normative data," *British Journal of Psychology*, vol. 74, no. 4, pp. 491–516, 1983.

[17] S. De Deyne, S. Verheyen, E. Ameel, W. Vanpaemel, M. J. Dry, W. Voorspoels, and G. Storms, "Exemplar by feature applicability matrices and other dutch normative data for semantic concepts." *Behavior research methods*, vol. 40, no. 4, pp. 1030–48, Nov 2008.